DOCUMENT RESUME

ED 323 068 RC 017 726

AUTHOR Kalsbeek, William D.; Parker, Rebecca Robin
TITLE Design Study of Methods for Sampling Migrant and Seasonal Farm Workers. Final Report.
INSTITUTION North Carolina Univ., Chapel Hill. School of Public Health.
SPONS AGENCY Milbank Memorial Fund, New York, N. Y.
PUB DATE 14 Jul 89
NOTE 137p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC06 Plus Postage.
DESCRIPTORS Agricultural Laborers; Farm Labor; Migrant Housing; *Migrant Workers; *Migration Patterns; Public Health; *Research Design; Research Methodology; Research Problems; *Sampling; *Seasonal Laborers; Social Science Research; Statistical Surveys

ABSTRACT

This report describes efforts to develop sampling methods to be used in national or regional studies of migrant and seasonal farm workers (MSFWs). Several facets of the MSFWs' lifestyle create sampling difficulties. One is mobility. Although the dynamic nature of MSFWs' movement is partly understood, it is sufficiently unpredictable to create problems, such as "multiplicity"--multiple opportunities for the same individual to be selected. Other problematic aspects relate to identifying all housing locations of migrants in a given area, the transitional periods between migrants' jobs, variations in the timing of growing seasons around the country, and the elusiveness of undocumented aliens. The proposed design involves a stratified four-stage cluster sample. The first stage uses area sampling methods for small groups of neighboring counties, while other stages use methods of list sampling for enumerating districts or block groups, migrant camps, and households. Sampling units, stratification, selection, and methodological issues are discussed for each stage. A large part of this document consists of Rebecca Robin Parker's 1989 master's thesis on this topic, which provides details on: (1) important migrant worker datasets compiled by the Migrant Student Record Transfer System, Douglas Massey, and Richard Jones; (2) definitions and characteristics of migrant farmworkers population; (3) multiplicity and how to handle it; (4) migration patterns; and (5) a description of a proposed 4-stage sampling design. An appendix contains a related paper by William Kalsbeek that discusses conceptual framework and design parameters and provides statistical formulas. (SV)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

William D. Kalsbeek

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

1. INTRODUCTION

The plight of the migrant and seasonal farm worker (MSFW) remains one of the significant yet largely overlooked public health problems in America. Recent studies paint a rather sobering picture of life among this small but growing segment of the population. Poor, uneducated, unskilled, and often illegal residents, their treks from town to town in search of employment can cause them to be found in virtually every state, and often great distances from their families and homes in Mexico, the Caribbean, or Central America. Strong economic motivations make them willing to work for low wages and live under substandard conditions to avoid further deprivation.

These living conditions, to which both the workers and their families (if present) are subjected, makes them vulnerable to a wide variety of both acute and chronic illnesses, thus creating health problems for themselves and those with whom they come in contact. The Migrant Health Act of 1962 was the first major piece of public legislation created to address these problems through the development of health care programs. While the programs emanating from this legislation have generally improved the migrants' quality of life, there still remains the need for researchers to establish a clearer understanding of migrant health problems through broad-based field research studies. In addition to some obvious problems with measurement and nonresponse, the matter of finding valid yet effective means to sample this elusive segment of the population has become one of the largely uncharted frontiers in the survey methods field and was the general goal of this study, which was conducted between June 1, 1987 and June 30, 1989.

A three-pronged strategy, similar to that used in developing most sampling designs, was employed to meet the study's overriding goal. First, we set out to better understand the lifestyle of the MSFW by examining the work of several of those who either study or serve migrants. Second, we identified these issues we thought would be most important in developing a regional or national sample of MSFWs. Third, we developed a preliminary working design that was modified as the key design issues were studied and our findings shared with colleagues in the service and research communities. We have concluded the following from our work:

> **There is no doubt that obtaining a scientifically defensible regional or national sample of MSFWs <u>is</u> plausible; however, implementing the process of sample selection would be relatively complicated and expensive.**

The scientific work on this grant was performed by two persons. The Principal Investigator, Dr. William Kalsbeek,

set the direction for the project as a whole and for the design development in particular. He was responsible for the methodological work done and did most of the information-gathering from colleagues outside of the University of North Carolina. A graduate research assistant in the Department of Biostatistics' masters degree program, Ms. Rebecca Parker, helped with the literature review done early in the study and did the statistical analysis of the MSFW mobility data. She consolidated much of the design issues information into what eventually became her masters thesis (see Appendix A).

Our format for presentation will be simple and brief. The process we followed to arrive at the design we recommend will be described in the next section. The structure of the design itself will be portrayed in the concluding section. In the interest of brevity only the highlights of our work will be reviewed. The reader interested in more detail will be directed to several appendices.

2. METHOD

In this section we discuss the process that was followed in meeting the goals set forth in the study. This process consisted of a series of steps which are typically taken in developing a design of the type mentioned above. The first several steps are to gather background information about MSFWs and from this information to develop a provisional design. Meanwhile several design issues are identified and addressed in some intermediate steps, and the provisional design is then modified to become the design that is ultimately put forth for further consideration.

Despite the explicit directive to examine ways to sample MSFWs, our work was affected adversely in two important respects. First, we were examining design issues and developing a statistical design in the absence of a real project on the immediate horizon. In most design settings, there are some explicit substantive goals for the survey which drive design development. There were none spelled out here, except that we could presume that analysis from the survey for which we were designing a sample would be aimed at profiling various health indices for MSFWs. Specifications were lacking on the measurements to be made, the population parameters to be estimated, and the population subgroups for which the estimates would be made. Second, the amount of money and the type of resources that would be available to conduct the MSFW survey was largely unknown. Reasoned speculation was our only recourse. It did become apparent, however, that any serious attempt at doing a comprehensive regional or national survey of this population would have to be done by an experienced survey research organization. We also reasoned that because of its operational complexity, a survey of this magnitude would

have to be supported by more than one of the potential sources of funding. Without knowing the precise level of funding for the survey we were unable to specify sample sizes at any of the sampling stages for our proposed design. These two limitations forced the final recommendations from our study to be aimed the structure of a design only.

2.1 Review Past Efforts to Sample MSFWs

We began by doing an extensive search of the literature for sample surveys where various segments of the MSFW population were the target of study. This effort turned up dozens of papers covering all aspects of the migrant lifestyle, from economic factors dictating their patterns of movement to their experiences in utilizing services provided for them. This extensive list of studies was screened for those where probability sampling methods were used to identify those individuals who were the subjects of study.

Existing sample surveys of MSFWs had a number of common characteristics. First, they were almost all local efforts, either covering a few neighboring counties or the clientele of migrant health clinics or other such migrant-directed service entities. A study aimed at several migrant locations in Colorado was the only state-wide effort we found. Some studies were aimed at communities in Mexico and other countries from which the various migrant streams originate. Second, these studies were almost always cross-sectional in nature. Time frames were short in duration in most studies, no doubt in the interest of minimizing the effect of mobility on the sample. Third, sampling designs were almost always relatively simple, reflecting their local target populations and the limits placed on the sampling operation brought about by the relatively labor-intensive process of producing lists of migrant housing for sample selection.

Our early review of the literature also led us to books, papers and monographs which helped us to better understand the migrant lifestyle and living conditions. Of particular interest was information which would enable us to better appreciate the dynamic nature of MSFW mobility and housing. Not only did we want to know where migrants might be found but when and where they go in the course of a year. This information would help us decide how to pick a sample of local areas in the first stage of the national design. In addition we needed to know the composition of the migrant household, specifically whether they tended to travel as individuals, in family groups, as groups of unrelated individuals, or as some combination of all three types. Where MSFWs live and the types of housing they live in were also important to be able to plan strategies for sampling within local areas.

We learned several things from this initial literature review. First, our early suspicions that MSFW mobility would be a key concern in formulating a sampling design were confirmed. MSFW movement is tied to economic, climatogical and social factors of their environment, which itself is somewhat unpredictable (e.g., rainfall and the demand for farm workers). Design development therefore could not be based on assumptions of random movement, nor could we presume that the patterns of movement would be totally predictable. Second, we discovered that defining the targeted population for any regional or national survey could be a debatable matter. Not only is there the basic distinction between migrant farm workers (those who temporarily move into an area to do work) and seasonal farm workers (those who live in an area year-round and who may do other kinds of work during the off-season), but there are also some notable differences in how these categories of workers are defined by various governmental agencies. Even once agreement is reached on the target group, ot was our impression that operational implementation of the definition could be difficult.

2.2 Contact Workers and Researchers in the Field of Migrant Affairs

After looking at much of the migrant research literature, we were able to identify some of those whom we thought would play major roles in any future survey of the migrant population and who could thereby provide helpful insight into how we should proceed. Many of these individuals were contacted by mail or by telephone to explain what we were attempting to do in our work and to solicit information which would be helpful in our subsequent work. (It was our belief that making these contacts early in the study would be important to avoid the trap of moving forward with sampling methods that ultimately would be found to be implausible.)

Several of these initial contacts became key sources of information and advice throughout the study. Included among them were Ms. Sonja Leon-Reig, Director of the Office of Migrant Health, who helped to identify useful data and information sources and provided a global perspective on migrant problems; Dr. Alice Larson, a Seattle-based private consultant in migrant affairs, and Dr. Carla Littlefield, a private consultant in the Denver area, who through prior research brought to the project a wealth of practical experience in sampling MSFWs; and Dr. Douglas Massey, an economist at the Center for Migrant Studies at the University of Chicago, and Dr. Richard Jones, a geographer at the University of Texas at San Antonio, whose data from community surveys in Mexico provided valuable insights into migrant mobility.

2.3 Identify the Major Statistical Issues to be Faced in Developing a MSFW Sampling Design

Several facets of the MSFW's lifestyle contribute to difficulties with sampling. One is their mobility. This segment of the population, variously estimated to consist of from 2-12 million persons, lives a nomadic existence, working in one location until the work is done or until they have accumulated sufficient resources to move on to the next location. This characteristic of the migrant lifestyle gives the sampling frame, the list from which the sample is drawn, a dynamic rather than the preferred static quality needed to simplify selection. Although the dynamic nature of this movement is partly understood, it is sufficiently unpredictable to create problems for the sampler.

The nomadic movement of migrant farm workers creates a statistical problem known as "multiplicity," in which a migrant can be linked to more than one unit in the sampling frame, the list from which the sample is chosen. For example, when a list of migrant camps is used and the period of the study is several weeks, it is possible by movement from one camp to the next that a migrant would have more than one chance of being selected. Without handling these multiple opportunities for selction in some reasonable manner, estimates from a sample would be biased.

Multiplicity has historically been handled in either of two ways. One is to establish a "unique counting rule" which, in effect, links each migrant to one and only one sampling unit on the frame. This solution unfortunately reduces the efficiency of the data gathering operation, because migrants will be screened out of the sample and the survey if they are not chosen through the sampling unit to which they are linked. The other solution involves determining how many opportunities the migrant had for selection and suitably weighting the data to compensate for the multiplicity. Whereas this remedy can produce unbiased estimates, the variation in sampling probabilities leads to somewhat larger variances of estimates (i.e., lower precision). Both solutions therefore have side-effects which would be particularly troublesome in sampling MSFWs since data collection, if done over a long interval, would lead to significant variation among multiplicities and thus amplify the limitations of these two approaches to the mobility problem. We felt as we considered the matter of dealing with multiplicity that perhaps another solution to this issue would have to be found.

A second problematic aspect of the MSFW's lifestyle when trying to sample them is that their place of residence may be difficult to locate and list during frame construction (i.e., the process of producing the list or lists from which a sample is randomly drawn). When this

concern is real, frame construction becomes both difficult and costly. If mishandled these difficulties translate into a statistical problem called coverage error, which can lead to survey estimates that are both less precise and statistically biased. MSFW camps provided by the employer can be readily identified in sampling housing locations within a local area, and MSFW households can usually be listed within most housing structures of this type, provided of course that the employer will allow such listing to be done.

The real difficulty comes in identifying all such housing locations within a local area, since MSFWs may be found in other less conspicuous locations like low-rent hotels, public housing projects, interspersed among other low-income private homes, or in remote camping areas. Unfortunately, the locations of many of these housing locations will not be known, even to some informed individuals in the local area itself. Obtaining a reasonably complete list of housing locations for MSFWs within a selected area of study would therefore require the time and funds to piece together a frame of locations from multiple sources in the community. Outreach workers from local organizations like the migrant health clinic, the health department, the housing authority, and legal aid would need to be contacted during what would be a lengthy frame construction process.

A further complication of the coverage issue is the high turnover rate in jobs held by migrants. Thus, at any given time there are some workers who are in transition between jobs or attending to personal needs, and therefore not included on the frame. The transitional period between jobs further affects coverage error in a sample of migrants and requires at minimum that the proportion of the migrant population be assessed and preferably as well that some means for including them on the frame be found.

Variation in the timing of growing seasons around the country creates another issue as to what would be an appropriate time frame for data collection in a regional or national study. Defining as short a data collection period as resources will allow and then interviewing MSFWs where they are living during that period would potentially bias estimates derived from measures that are tied to location (e.g., access to housing and health care services). On the other hand, lengthening the data collection period increases the likelihood of migrants appearing in multiple locations.

2.4 Develop a Provisional Design

Once we developed a working understanding of life as a MSFW and thereby caught a glimpse of the issues that would face us in sampling this population, a provisional design

for a regional or national study could be produced. This first draft of the main product in this study saw us envisioning a stratified four-stage cluster design which employed many of the common features of large-scale designs in national surveys today. For example, sampling units in the first two stages were county groups and enumeration districts/block groups, area units often used in household surveys of the general population. As the issues which accompanied the statement of the design attest, there were yet some final features to be developed which would accommodate some unresolved statistical issues at that point.

A written sketch of the provisional design (see Appendix B) was prepared and circulated for comment to several statistical colleagues as well as those in the migrant services community. The provisional design therefore became the culmination of our early thinking and formed a basis for further discussion and development work.

2.5 Solicit Reaction to the Provisional Design from the Research and Service Community

With a working prototype for the sampling design now in hand we were ready to share our ideas with statistical colleagues as well as those in government and the health care community who would become the users of data gathered from a survey of MSFWs. This two-pronged approach to evaluating the provisional design was thought to be necessary to assure both the scientific merit and practical utility of our work.

In addition to the solicitation of comments from selected members of the statistical community at large, we submitted our ideas to the program planning committee for the 1988 national meeting of the American Statistical Association held in New Orleans. Our paper was accepted and placed in an invited session that was devoted exclusively to the matter of sampling elusive populations like MSFWs. The response to the paper was largely positive and several comments received at that time helped us to better focus our remaining work.

Exposure of our ideas (at various stages of development) to persons who routinely work with migrants was accomplished by taking three field trips to areas with relatively high concentrations of migrant labor and by making presentations at two migrant conferences. We sought through the trips: (1) to share our ideas with people who would be the best judges of the practicality of our ideas, (2) to immerse ourselves directly into the migrant culture to thereby better understand the people we would be sampling, and (3) to observe the various locations and types of housing in which MSFWs are found. The migrant conferences provided us with forums: (1) to share our

perspective of the problem of sampling MSFWs, (2) to hear reaction from potential users of our ideas, (3) to solicit needed information for our designs work concerning the patterns of enrollment in migrant camps, and (4) to discuss various current efforts around the country by workers in migrant health clinics, health departments and other such agencies to study migrants.

The field trips were geographically diverse so that we could gain the broadest possible perspective on the MSFW's living arrangements, which we knew would vary around the country. The first trip was to a migrant health clinic located in central North Carolina and about 50 miles from Chapel Hill. The second trip was to several migrant clinics in Northern California near Stockton and in the Yakima Valley of Central Washington. The last trip was to several migrant farming areas near Boulder, Colorado.

We presented our sampling ideas for discussion at two conferences specifically aimed at the health care needs of MSFWs. One presentation was made at a plenary session of the East Coast Migrant Stream Forum in Ashville, North Carolina. The other was for a similar session at the 12th national Migrant Health Conference in Indianapolis, Indiana.

2.6 Find a Workable Remedy to the Multiplicity Problem

During the time period when the field trips were being taken and the presentation made, we were also studying the statistical problem of multiplicity that is created by the unique mobility patterns for MSFWs. The novel solution we have found and shared particularly with colleagues at the 1988 ASA meeting is to sample jointly in space and time. This approach is unique to the survey research community, which creates spatial designs for most of its surveys, even those done longitudinally, and thereby generally disregards efforts to sample along the temporal dimension. The principal advantage of the time and space application to the problem of sampling MSFWs is that the multiplicity problem can be averted altogether.

A copy of the manuscript which more thoroughly describes our solution to the multiplicity problem is presented in Appendix C. A shorter version of this paper will appear in the Proceeding of the Section on Survey Research Methods from the 1988 ASA meeting. The presented version is currently being revised for publication in Survey Methodology.

2.7 Analyze Data on Mobility

Our early communications with those doing migrant-related research enabled us to identify and obtain

three data sources that we thought would shed some much-needed light on the matter of migrant mobility. Two were from sample surveys of migrants to the U.S. in several Mexican communities. These data had been gathered in separate studies conducted by Drs. Douglas Massey and Richard Jones (mentioned previously in this report). The third source of data was the Migrant Student Record Transfer (MSRT), an information system which has been developed and maintained through funding from the Office of Migrant Education. The MSRT contains information on movement by families with enrolled school-age children. Because of its size only a sample of the MSRT file was obtained for our analysis.

These data sources helped us address two important questions whose answers would have an impact on our final solution to the matter of dealing with MSFW mobility. One question was how far MSFW households move when they move from one job to the next, and the other was how long do they stay when they find work. Findings on both of these are summarized as part of Rebecca Parker's Master's thesis found in Appendix A.

2.8 Prepare Final Design Recommendations

The final step of the study was to revise the preliminary design in accordance with subsequent work and commentary from persons external to the project. Insight and information provided from presentations, trips, and analysis of data were all used in arriving at the final design product of this study. It is our sincere hope that our initial efforts at finding suitable ways to sample MSFWs will be subject to further scrutiny and revision, where necessary, and that our beginning work with this problem will help pave the way for a regional or national MSFW survey that meets the growing information needs for this segment of the American population.

3. SKETCH OF DESIGN RECOMMENDATIONS

The sample for a regional or national survey of MSFWs could be selected by means of a design which involves selection both in space and time. The spatial-dimension sample of MSFW dwellings might be chosen by means of a stratified, multi-stage cluster design. The time-dimension of the design would be implemented by choosing a sample of time segments (e.g., individual days) for each dwelling units chosen in the spatial sample. MSFWs occupying the sample of dwellings during the selected time segments would be interviewed. Selecting this sample would utilize standard methods of area sampling through the first stage of selection and then of list sampling in the remaining stages.

General issues: (1) Which of several definitions of MSFWs adopted by the various federal agencies and groups would be used for operational purposes, and would the definition include the workers only or their families as well? Where do U.S. citizens who do this type of work fit in, since they are subject to most of the same health problems as those who cross international boundaries to work in American fields

(2) What would be the funding source(s) for this national survey and what would be their research agenda for the study?

(3) What would be the funding level for the survey?

(4) Would the study be conducted by an established survey research organization or a coordinated effort of a large number of more local groups (e.g., migrant health clinics, state health departments, county health departments)?

First Stage

Sampling unit = Small group (i.e., 3-6) of neighboring counties

Stratification = 3-5 migratory streams; concentration of MSFWs according to the Department of Labor; other geographic predictors of health status

Selection method = Probability Proportional to Size (PPS) selection with estimated average number of MSFWs during the period of data collection as the measure of size; size measure from the most recent census or other reliable source acceptable as long as measures are available at the county level

Issues: (1) How many counties and how large an area should be covered by the counties grouped together as PSUs; larger groups tend to reduce the chances of double-counting especially mobile MSFWs, while smaller groups would reduce the cost of field operations for sampling and interviewing

(2) Considerable cost savings could be realized by excluding counties with low concentrations of MSFWs; the drawback to this would be a failure to achieve complete coverage since some MSFWs would

exist in the excluded areas

(3) This stage of sampling depends heavily on the existence and availability of accurate county-level census (of population or agriculture) data on the number of MSFWs or the number of households they occupy; although this information on MSFWs has been collected in recent censuses, these aggregated figures are not routinely produced and would require special arrangements with the Bureau of Census and/or USDA; because of other reporting priorities, there might be some difficulty in getting the needed figures on a timely basis; thus, anyone following this design for a MSFW survey in the early 1990's and relying on census data would face the following choices: (a) use 1980 figures on the number of MSFWs for sampling purposes, (b) wait for the 1990 figures and face the possibility of delays in the survey, (c) use some other size measure for PPS selection if the needed figures cannot be easily produced

(4) Construction of the size measure would also require knowledge of the local growing seasons for each PSU, in order to anticipate local fluctuations in the MSFW population

Second Stage

Sampling Unit = Camps and other specifically identifed clusters of MSFW housing (e.g., motels, housing projects, neighborhoods)

Stratification = Form two substrata: (1) housing clusters intended mainly for seasonals, and (2) housing intended mainly for migrants

Selection = Select the same number (b) of housing clusters from each PSU in substratum (1) with PPS using the best local figures on the number of occupants as the size measure

Issues:

(1) b would be determined later on the basis of what is thought to be an optimum allocation of the sample within PSUs

(2) This stage relies heavily on our ability to identify and obtain the measures of size for MSFW housing at the local level; this, however, has been done

successfully in prior surveys although at great cost and with considerable effort expended in order to produce a sampling frame with complete coverage; frame construction at this stage will require help from those involved in outreach activities for local public and private groups providing service or assistance to MSFWs (e.g., migrant health centers funded by the USPHS Office of Migrant Health, housing agencies, legal aid societies, church-affiliated relief organizations, migrant advocacy groups, local health departments, agricultural extension offices); information from several of these sources would have to be merged to produce the final frame

Third Stage

Sampling unit = Migrant dwelling unit (i.e., with a "dwelling unit" as defined by the Bureau of Census or an individual within a "group quarters" as defined by the Bureau)

Stratification = Type of migrant housing unit; form the following two substrata: (1) dwelling units and (2) individuals in group quarters; substratum (1) would tend to contain MSFWs and their families, while substratum (2) would have a high percentage of single men living in houses and dormitories

Selection = Systematic sampling from a concatenated list of the MSFW dwelling units in the two substrata

Issues:

(1) Frame construction may be a challenging and labor-intensive task in this stage, requiring arrangements to access some potentially difficult places and high-risk areas (e.g., migrant camps located on grower's property, public housing project)

(2) Because of the amount of effort likely to be needed for frame construction, perhaps it would be more cost-effective to have the field work for frame construction and the interviewing done at the same time and maybe even by the

same person; traditionally this combination of activities has been avoided because of the potential for invalidated sample selection caused by field workers who abandon the random methods in favor of sampling for the sake of convenience rather than science; of course frame could be sent to a supervisor or a central location for selection, but that may contribute to delays

Time-Dimension Sampling Applied to Selected MSFW Dwellings

The description of the sampling design thus far covers the spatial dimension of the sampling design through three stages of selection. A related segment of the design to deal with the problem of MSFW mobility involves sampling in time. Precisely how sampling in this dimension is carried out would depend on the nature of the data collection protocol for the survey (in particular, the questionnaire) and the length of the data collection period. Concerning the latter, it is thought that for a national survey of MSFWs it would be best if budgets were to allow to collect data over a 12 month period, to deal with issue of seasonality for many important MSFW measures (e.g., health risk).

Sampling Unit: Time segments (e.g., a single day or a small number of consecutive days, depending on the data collection plan)

Selection: Systematic sampling from a list of MSFW dwellings ordered over the period of data collection; a separate time sample would be chosen for each MSFW dwelling in the spatial sample

Issues:

(1) One of the practical problems with randomized time sampling separately for each selected MSFW dwelling is the likely loss in operational efficiency; it may be easier and less expensive to select separate time samples within SSUs or PSUs (i.e., the same days would be sampled for each sample dwelling within these sampling units); the feasibility of this revision would partly depend on sample sizes within PSUs and on the work force used for interviewing

(2) Scheduling could be made difficult since data would have to be collected on the selected days; the use of multiple-day

time segments as sampling units would help in this regard

Fourth Stage

Sampling unit = Individual MSFW living in a dwelling during one of its selected time segments

Stratification = None currently planned, although the ordering by age for systematic sampling constitutes an implicit form of stratification;

Selection = Systematic sampling from an age-ordered roster of persons listed in the MSFW dwelling unit; depending on the nature of the survey questions, it may be advisable to interview only one person chosen at random in the selected dwellings

Issues:

(1) Explicit stratification would be needed if the decision is made to oversample certain segments of the MSFW population (e.g., Haitians, females, legal MSFWs, etc.); the challenge then would be to find a suitable strategy for having the interviewer validly apply a randomized method with higher rates for the oversampled group(s)

(2) This stage of sampling would only be needed if, because of the type of data to be collected, it is necessary to select one person at random in each migrant housing unit (e.g., opinions are often contaminated if multiple interviews are conducted in a household, thus making a single interview per household desirable); in the event that one interview per household is not necessary, the logical choice for most types of data is to interview all eligible members in each sample household

APPENDIX A

Design Issues in Sampling Migrant Farm Workers

A DESIGN STUDY ON THE PRIMARY PROBLEMS OF IMPLEMENTING A NATIONWIDE STUDY OF MIGRANT FARMWORKERS IN THE US

by

Rebecca Robin Parker

A paper submitted to the faculty
of the University of North Carolina
in partial fulfillment of the requirements
for the degreee of Master of Science
in the Department of Biostatistics

Chapel Hill

1989

Approved by:
William Kalsbeek, Ph.D.
Adviser

Gary Koch, Ph.D.
Reader

Craig Turnbull, Ph.D.
Reader

To my very relieved parents, Sieg and Sylvia Parker,
my patient advisor, Dr. William Kalsbeek,
Keith Brown, for his computer expertise,
& Marshall Rohde- just for being there
when I needed him.

TABLE OF CONTENTS

	PAGE
I. Introduction	
A. Goals of a Nationwide Study	1-2
B. Quality of Life- Living and Working Conditions	2-3
C. Issues of a Design Study	3-4
D. Descriptions of Datasets	
i. Migrant Student Record Transfer System	4
ii. Massey	5
iii. Jones	5
iv. Similarities of Jones and Massey datasets	6
II. Defining the Migrant Farmworker Population	
A. Examples of Migrant Farmworker Definitions used by Different Sources	7
B. What is a Migrant Farmworker?	
i. Distinctions between Migrant Seasonal Farmworkers and Migrant Farmworkers	8
ii. Definition of the Term "Migrant"	8-9
iii. Definition of Farmworker	9-10
iv. Age Limitations	10
v. Suggestion for Definition of Migrant Farmworker	10-11
III. Migrant Farmworkers are a Dynamic Population	
A. Illustrating this Point with Richard Jones Literature	12
B. Examining Occupational Trends in the Jones and Massey Datasets	13-14
IV. Multiplicity	
A. Definition of Multiplicity and How to Calculate It	15-16
B. Weighting versus Unique Counting Rule	16
C. How Local Studies Dealt with Multiplicity	16-17
V. Mobility	
A. Migratory Streams are Defined	18-19
B. What motivates Migrant Farmworkers to move?	
i. Agricultural Seasons	20
ii. Weather	20
iii. Legal Status	20-21
VI. In Depth Discussion of Migration Patterns	
A. How Often do Migrant Farmworkers Move?	22-23
B. How Far do Migrant Farmworkers Move?	23-25
C. Family versus Solitary Travelers	25-26

	PAGE
VII. Difficulties in Finding Migrant Farmworker Housing	27-28
VIII. Proposed Four Stage Design	
A. Stage 1- County Groups	
i. Should Counties with Low Concentrations of Migrant Farmworkers be Included in the Survey?	29-30
ii. How to Determine Migrant Farmworker Population Size to Serve as Measures of Size for PSU Selection	
a. Department of Labor	30-31
b. Census Bureau	31
c. US Department of Agriculture	32
iii. Composition of the Primary Sampling Unit	32
B. Stage 2- Enumeration Districts/ Block Groups	32-33
C. Stage 3- Migrant Camps	33-35
D. Stage 4- Migrant Housing Unit	35-37
IX. Conclusion	38-40
Appendix I	41-66
Appendix II	67-72
Bibliography	73-74

A DESIGN STUDY ON THE PRIMARY PROBLEMS OF IMPLEMENTING A NATIONWIDE STUDY OF MIGRANT FARMWORKERS IN THE US

BY

REBECCA ROBIN PARKER

I. INTRODUCTION

This design study, sponsored by the Milbank Memorial Fund, identifies and discusses the major issues encountered when sampling a nomadic population- namely migrant farmworkers in the United States. Since a nationwide study of migrant farmworkers has never been attempted, the documentation on such a wide undertaking is scarce. Information was gathered by examining local studies, talking with individuals who have studied and/or sampled migrant farmworkers, and analyzing secondary data sources. Making use of these resources, I will identify, address, and make several reccommendations concerning several design issues encountered in a nationwide study. In addition to discussing issues, a preliminary 4-stage design is suggested. Keeping in mind that the issues raised in this study do not always have obvious solutions, the contents of this paper are intended to serve as a foundation from which to begin a nationwide study.

A. Goals of a Nationwide Study

Before examining the logistics involved in sampling migrant farmworkers in the US, it is important to briefly establish why a national survey on migrant farmworkers is necessary.

Migrant farmworkers are among the poorest working persons in our nation. The 1985 average family income ($6,367) of a migrant farmworker places them well below the poverty level. (Littlefield, Stout 1988) Across the US today, the poverty level for a family of 4 is $11,650. (Allison, 1988) With this poverty comes atrocious working and living conditions. Migrant farmworkers cannot form a union to protest working conditions since the they are exempt from the Taft-Hartley Act. (Johnston, 1985) They cannot search for alternate jobs since they are often uneducated and unskilled. If they originate from a foreign country, they may be unable to speak English and may even be in the US illegally. Thus, with no better options, they remain in an occupation which is dangerous, hazardous and exploitative. (Littlefield, Stout 1987) Sonia Leon-Reig, director of the Office of Migrant Health, speculated about the analytic goals of a national survey in her July, 1987 conversation with Dr. William Kalsbeek. She suggested the following goals : 1) Obtain a measure of the health status of migrant farmworkers in such areas as prevalence of tuberculosis, number of days spent in a hospital, and/or disability 2) describe occupational health, the conditions in the workplace 3) determine a profile of users of migrant health centers and 4) obtain an estimation of the number of migrant farmworkers. [Number 4 is given the lowest priority]. Obviously, a nationwide study will not solve the problems of migrant farmworkers. However, the scarcity of information on this group demands that a nationwide study be implemented. Unless conditions are more fully known and better understood, programs to help the migrant farmworkers cannot be implemented.

B. Quality of Life- Living and Working Conditions

Of all the employment groups in the US, farmworkers have the poorest physical and mental health. (Larson, 1982). Not incidentally, agriculture is reported as the second most dangerous occupation in the US. (Wilk, 1986) The migrant farmworkers' work environment is a major cause of health problems, especially when we consider exposure to pesticides.

dangerous farm machinery, unsanitary field conditions, and substandard housing. Few health and safety standards are set and even these are not adequately enforced. Agricultural machinary may lack proper safety features causing unnecessary accidents. Migrant farmworkers are exposed to toxic pesticides without sufficient protection. (Johnston, 1985) Many labor camps are located adjacent to fields which have been sprayed with dangerous pesticides. (Wilk, 1986) To make matters worse, some migrants may not even have housing and actually live in the pesticide sprayed fields in which they work. (Wilk, 1986, Larson, 1982) The fields in which they work often lack toilets and adequate drinking water. It was estimated in a 1984 analysis that only 22-45% of hand-labor-intensive farmwork across the US have management provided sanitary facilities and drinking water. (Wilk, 1986) Without these two essentials, migrant farmworkers are more susceptible to communicable diseases, skin rashes, heat disorders, and urinary tract and kidney infections. (Wilk, 1986) Migrant housing, which may be furnished by the farmer doing the hiring, is also grossly substandard. One study in Colorado even went so far as to label migrant housing which often lack basic sanitation, refrigeration, and hot water as "approximate [to] those of the Third World." (Littlefield, Stout 1987).

Obviously, from a humanitarian perspective, migrant farmworkers are a population in dire need of assistance. Migrant farmworkers suffer from parasitic infections, tuberculosis, childhood diabetes, anemia,low birth weight, heat stress, gum disease and work injuries more so than the general population. (Quillan, 1988) Hopefully, by pinpointing specific problems and areas of neglect, a nationwide survey would enable migrant health care providers (and others who work on a regular basis with migrant farmworkers) to be more effective in handling the needs of the migrant farmworker.

C. Issues of a Design Study in Brief

A design study applied to the population of migrant farmworkers in the U.S. requires

that several issues be resolved. These issues include: the problems of defining a migrant farmworker, multiplicity, geographic mobility, occupational mobility, and elusiveness. These terms may seem vague and inconsequential at the present time , but their significance will become apparent as they are separate topics of discussion in subsequent sections of this paper.

D. The MSRTS, Massey, and Jones datasets

Before discussing these design issues in depth, background information on the three datasets analyzed for this study will be briefly given. Further details on these datasets, as well as explanations for exhibits associated with them can be found in Appendix I and Appendix II.

i. MSRTS

The Migrant Student Record Transfer System (MSRTS) is a database system kept by the Department of Education which contains health and educational information for some 750,000 migrant children. (Schlegal, 1987) Analysis of the MSRTS dataset will help us understand the migration patterns of migrant farmworkers who have children enrolled in a Migrant Education Program. Although this database has certain shortcomings (see Appendix II) , it is the only migrant program with a national record keeping system. (Larson, 1982)

In order to initiate the study, we contacted the MSRTS in Little Rock, Arkansas and received permission to use their database. Since we were not allowed direct access to their database, we requested a random sample based on the following systematic list sampling technique. The tenth child listed in the database was chosen and every 50th child thereafter. [The MSRTS contains records for approximately 125,000 "currently migrating children" of MFWs in the years 1985-1987]. This technique resulted in a sample size of 2500 migrant children or roughly 2% of the migrant population.

ii. Massey

Douglas Massey, of the Center for Migration Studies, has studied the geographic mobility of Mexican migrants who travel to the US. We obtained a tape of the data he gathered in his Mexican Migration Project. This tape contains the work histories of people (mainly household heads) from two of the four sample communities Massey studied in Mexico. [The two communities are Altamira and Chamitlan.] Since we were only interested in US migrant farmworkers, an individual had to have been employed in agriculture in the US at least once to be included in our analysis. To be consistent with the Jones data (to be discussed later), labor histories between the years 1976-1983 were analyzed. Of the 616 life histories available on this dataset, only 109 contained records of US agricultural employment in the years specified.

iii. Jones

Richard Jones, a geographer who has studied the occupational and spatial mobility of Mexican migrants in the US, was a significant source of information in this study. After a review of Jones' literature, he was contacted and arrangements were made to obtain some of the data he had gathered for his research. Specifically, we received over 200 pilot questionnaires administered to Mexican migrants in Rio Grande, Zacetecas on January 1984. Since half of these questionnaires were in Spanish, a Spanish translator was necessary to help decipher the results.

The only information of immediate interest was the migrant's work history in Mexico and the US. Only jobs held between 1976 and 1983 were examined. The state the job was held in, the year of employment, and duration of employment were recorded. Distinctions were made between agricultural and non-agricultural jobs. If a questionnaire did not have a migrant employed in US agriculture between 1976 and 1983, it was not used. Only 96 questionnaires met these criteria and were included in analysis.

iv. Similarities of Jones and Massey Data

Both the Jones and Massey data are similar in that the communities of interest are located in Mexico. Since a high percentage of migrant farmworkers on the West Coast and mid-continent streams are Hispanic, their choice of subjects seems justifiable. In Texas, the main destination of Jones' study population, 95% of the migrants are Hispanic. (Okador, Richards, Slessinger,1982) As further evidence of the predominant presence of Hispanics, in a study orchestrated by by the Department of Health, Education, and Welfare (HEW) in the Washington, Oregon, and Idaho area, 89% of the migrants sampled were Hispanic. (Larson, 1982). Thus stated, Mexican migrants represent a significant proportion of the MFWs found in the US. Analyzing the Jones and Massey data will aid in our understanding of the migration patterns of migrant farmworkers in the western and midwestern US.

II. DEFINING THE TARGET POPULATION

Before a nationwide study can be attempted, the target population must be specifically defined. The design of the study and size of the target population should be suited to the particular definition chosen.

A. Examples of Migrant Farmworker Definitions used by Different Sources

SOURCE	DEFINITION OF MIGRANT FARMWORKER
US DEPT. OF AGRICULTURE -ECONOMIC RESEARCH SYSTEM	Migratory farmworkers are those who 1)left their home temporarily overnight to do hired farmwork in a different state with the expectation of eventually returning home, OR 2) had no usual place of residence, and did hired farmwork in 2 or more counties during the year (Larson, 1982)
US DEPT OF LABOR	Employed, domestic seasonally hired agricultural worker, age 16 or older, away from residence at least overnight to do farm wagework. (Schlenger, Ordrizek, Hallan, 1979)
US DEPT OF AGRICULTURE	Persons 14 or older who do at least one day of farm wagework per year and must leave home at least overnight to do farmwork (Schlenger, Ordrizek, Hallan, 1979)
DEPT OF HEALTH AND HUMAN SERVICES	An individual whose principal employment is in agriculture on a seasonal basis, who has been so employed within the last 24 months, and who establishes for the purposes of such a temporary abode. [This definition was used for the Migrant Health Act.]* (Johnston, 1985)
ALICE LARSON**	Migrant seasonal farmworkers provide temporary season farm labor, such as picking, thinning or weeding crops and fruits: and who work in food processing plants and in other areas of agriculture which require hand labor. (Larson, 1982)

*Migrant Health Act (1962) has the main purpose of improving the health of migrant farmworkers. This law is resposible for more than 300 sites of migrant health delivery. (Johnston. 1985)

**Alice Larson is the migrant coordinator for the Employment Standards Administration, US Dept. of Labor

This table illustrates the wide range of definitions which exist for migrant

farmworker.

In the discussion below, different aspects of these definitions will be examined.

B. What is a migrant farmworker?

i. Distinctions between migrant seasonal farmworkers and migrant farmworkers

The innocuous term 'migrant' needs to be specifically defined due to the different situations which abound for those engaged in agricultural work. In the literature, migrant farmworkers and seasonal farmworkers have frequently been combined into the term migrant seasonal farmworker. (MSFW) Unfortunately, this may form some confusion since migrant farmworkers and seasonal farmworkers form two very separate groups. All of the above definitions agree that to qualify as a migrant farmworker, one "cannot return to their normal abode at night". Seasonals, however, can return home at night. Seasonals would include such individuals as teachers or students who work on a farm during their summer months. Although they are engaged in agricultural work, seasonals are able to return to their own homes. Migrants, on the other hand, would usually live in housing furnished by their employers. Migrants would have migration patterns while seasonals remain for the most part stationary. (Larson, 1982) Although the distinction between migrants and seasonals seems readily apparent, it should be noted. This paper will focus only on migrant farmworkers and the acronym MFW will be used.

ii. Definition of the term 'migrant'

How literally the term 'migrant' should be defined is another issue. Some farmworkers may live in 'temporary abodes' in the US, but are nonmobile. For example, agricultural workers are regularly recruited from the Caribbeans via the H2 Plan. Briefly, the H2 plan is mainly for Caribbean migrants. 20,000 workers are admitted annually into the US to engage in seasonal agricultural work- 1/2 of these workers cut sugar cane in Florida. However, once they are in the US, they remain stationary and are forbidden to

leave their work in search of better employment. These individuals have no migratory patterns and all of their health care is provided by their employers. (McCoy, 1985)

Although 20,000 is not a substantial percentage of the total migrant farmworker population, the migratory issue is also pertinent in other situations. Some migrant farmworkers have chosen to 'settle out'. These individuals start out as migrant farmworkers, settle in a community, and continue in farmwork. They may even continue to migrate at a later date. Although, during the growing season, they may live in housing supplied by the growers, they would not be migrants according to the US Dept. of Agriculture- Economic Research System's definition. The USDA-ERS definition has a spatial as well as temporal dimension. To "qualify" as a MFW, one must move between states or counties. This issue arose after examining a dataset in our study from the Migrant Student Record Transfer System in which a significant percentage of migrants (33%) were labeled "settled out". These MFWs may be settling in one location while in the US and commuting to different job locations. Although, technically, they do not possess migration patterns , they are living in "temporary abodes." Unlike MSFWs, descibed earlier, those who have "settled out" are more apt to earn a major portion of their incomes from agricultural work than MSFWs who may have only a casual attachment to agricultural employment.

iii. Definition of Farmworker

The term 'farmworker' is very broad and can encompass many fields. According to the Public Health Service Act, Title III, Part D, the term "agriculture" means

> farming in all of its branches including
> (a) cultivation and tillage of the soil
> (b) the production, cultivation, growing, and harvesting of any commodity grown on, in, or as an adjunct to or part of a commodity grown in or on the land (Johnston, 1985)

In contrast, Alice Larson's definition includes agriculturally related non-farm jobs. [Included among farmworkers are those who work in food processing plants.] However, most

studies on MFWs seem to be limited to the farm itself. A decision should be made whether or not to place restrictions on the types of agricultural related work which are to be included in the study. For example, should those who work on ranches, which technically do not qualify as farms be included in a national survey?

iv. Age Limitations

Another issue is whether or not to place age limitations on being a migrant farmworker. For example, in the US Department of Labor (USDL) definition, an individual must be at least 16 before he is to be considered a migrant farmworker. The US Department of Agriculture (USDA), however, places its age limitations at 14. This matter should not be taken lightly since a considerable number of children may actually work as "hired hands". Under the Child Labor Laws in seasonal agriculture, children under 16 are allowed to work in the fields. [However, they are prohibited from the more hazardous tasks. (Johnston, 1985)] The decision of whether or not to include children as MFWs will specifically affect the 4th stage of the proposed design which involves the selection of individuals to be interviewed for the study. (See Section VIII.)

v. Suggestion for definition of a Migrant Farmworker

Although many definitions for MFW exist, the definition developed by the Dept. of Health and Human Services causes the most concern to us since it is the definition utilized by the Migrant Health Act. In this definition, a MFW is one who at some point during the past 24 months has worked in agriculture on a seasonal basis. If this definition is to be taken literally, at the the time of the actual survey an individual could "technically" be a MFW if he worked in agriculture fifteen months ago (and fulfilled the other qualifications to be classified as a MFW.) Perhaps, this individual has permanently moved out of this field of work and is no longer working in agriculture. Thus, additional effort and expense would

be necessary to include these individuals in the survey. If we adhered to this definition, a dual frame, one aimed at those currently working in agriculture and another for those who have moved on to other nonagricultural work, would be required to prevent undercoverage.

One possible solution to this problem is to define as a migrant farmworker those who are migrant farmworkers at the time of the actual survey. A migrant farmworker is not necessarily a migrant farmworker for a continuous length of time. There seems to be a very high degree of job turnover.

AGRICULTURAL JOB ---> NON-AGRICULTURAL JOB ---> AGRICULTURAL JOB

Individuals who are not MFWs from Jan-March may be MFWs later in the year (and vice versa.) Thus, extending the period of study over a year's duration would mean a higher coverage of MFWs. A year long study would also be beneficial since it allows us to examine the working and living conditions of MFWs in different environmental conditions. The health needs of a MFW in the winter will no doubt be different from his health needs in the summer.

III. MFWs ARE A DYNAMIC POPULATION

A. Illustrating this point via Richard Jones literature

The high degree of migrant turnover which is discussed in the preceeding paragraph can be shown in the occupational mobility studies of Robert Jones.(1984, 1986) Jones did two separate studies on Mexican migrant workers in the South Texas region. One study (Jones,Murray,1986) compares the employment records of migrants from two different Mexican communities- Zacetas [located in the interior of Mexico] and Coahuila [located on the border of Mexico.] Unfortunately, Jones does not have complete employment records for the migrants in this study. Information is limited to the first job in the US, the first urban job in the US, and the latest job in the US.

The results of this study illustrate the temporary nature of farmwork. In Zacetas, 50% of migrants had their first jobs in agriculture or sawmills. [Jones grouped these two categories together.] By their latest job, only 30% of migrants were agricultural/sawmill workers. The migrants from Coahuila showed a similar trend in job turnover. 36.4% had first jobs in agriculture while 27.3% remained in agriculture for their last job.

A second study (Jones,Harris,Valdez, 1984) also illustrates a high degree of occupational mobility. Two thirds of the migrants' first job in the US were agricultural while, by the latest job, only one third were agricultural. Interestingly enough, one third of those migrants who held at least one urban job [construction, service jobs] prior to their latest job went from the urban sector to the "agricultural sector." This shows that movement between agricultural and non-agricultural jobs is not always in one direction.

B. Examining occupational transitions in the Jones and Massey datasets

Before a discussion of the Jones and Massey datasets commences, it is important to briefly describe the basic composition of the exhibits (in Appendix I) used to explain my findings. The geographic and employment patterns of MFWs are separated into two distinct categories- namely "within-year transitions" and "between-year transitions". Within-year transitions exist for MFWs who held more than 1 job within the year. Between-year transitions compare the last job of a year with the job of the succeeding year. Both transitions are relevant since we are concerned about short-term movements made within a year and the patterns of change over a longer duration of time. Using the calendar year as the basic unit of measurement is the logical choice due to the cyclical nature of the growing seasons. Also note, in the Jones and Massey exhibits, employment records are separated into "agricultural" and "nonagricultural." An agricultural job, in this context, implies that the work was done while migrants were residing in the U.S.

The Massey and Jones datasets also demonstrate the dynamic nature of the MFW population. Many MFWs do not remain in agriculture related jobs on a continual basis. Examining Jones exhibits 1-3 in Appendix I, one observes that although 60.7% of all moves within the year were within US borders, only 19.7% of all moves within the year were from an agriculturally related job to another agriculturally related job. Thus, only 32.4% of all within-year moves confined to the US involved agricultural jobs exclusively. Between-year transitions showed very comparable trends. While 58.2% of all moves for between-year transitions were made within the US, only 23.7% of all moves involved only agriculturally related jobs. Thus, only 40.7% of all between-year movements restricted to the US were between agricultural jobs. (See exhibits 4-6.)

The Massey dataset, however, reveals much less diversity in the choice of occupation. Between-year movements illustrate that most of the work done in the US was agriculturally

related. While 24.8% (see exhibits 12-14) of all between-year movements are restricted within US borders, 18.6% of all between-year movements are from an agriculturally related job to an agriculturally related job. Hence, 75% of all between-year movements confined to the US are exclusively agricultural. In fact, of all the work done in the US, 86% was agriculturally connected. [Within-year movements were not referred to since only one MFW held more than one agricultural job in a year. This may be explained by our not having a complete within-year work history from the Massey data.]

The Jones and Massey data do show some similar mobility trends as well. Although the Massey MFWs do not show diversity of occupation while in the US, they do exhibit constant and frequent movements between the US and Mexico. For movements made within the year in the Massey dataset, 73.1% were made between Mexico and the US. (See exhibits 10, 11) In between-year movements, 26.4% are made between the US and Mexico. (See exhibits 13, 14) In the Jones dataset, approximately 38% of movements made between-years and within-years are international. (See exhibits 3 and 5.)

The point is that MFWs are not necessarily employed in agriculture on a continual basis. They may switch the type of work they do or move back to their homeland. Thus, a year long study has the distinct advantage of improved sample coverage of whatever becomes the targetted population. The main drawback of this strategy, however, is that now all of this movement which has been profiled will make sampling more complicated. Mobility will give some MFWs multiple chances for selection and create the issue of multiplicity which is discussed next.

IV. MULTIPLICITY

A. Definition of multiplicity and how to calculate it for a MFW

Now, since ideally, the study would take place over a year's duration and would cover the entire nation, multiplicity becomes an important issue. Basically, multiplicity occurs when a member of a population has a multiple probability of being selected. For example, a migrant farmworker who goes to five different locations (and thus five different migrant camps) in the time span of a year [the identical time span of the survey] is said to have a multiplicity of 5. Theoretically, this individual could be included in the sample five times. A migrant farmworker who moves only twice in a year would have a multiplicity of 2. Obviously, a migrant farmworker who is very mobile has a higher multiplicity than his more stationary counterpart. To determine the multiplicity of an individual migrant farmworker, we need to know how often this MFW moves in a timespan of a year. If the survey begins in January and we interview a MFW in January, the only method to determine multiplicity is to ask one of these two questions. 1) How often do you plan to move in the course of a year or 2) How many times have you moved in the last 12 months? [*Note: A 'move' occurs when a MFW changes the location of his residency. However, changing location of workplace while remaining the the same residency does not constitute a "move." The MFW must also be currently employed as a MFW while living in this dwelling.] Question 2 is preferable since this is based on the actual experience of the MFW. Although it may not specifically apply to the period of the study, asking the MFW how many times he plans to move is very subjective and is difficult for the MFW to predict. [A MFW has no control over economic conditions, weather, etc.] Since migrants have different probabilities of being

included in the sample, appropriate weighting is necessary.

[SAMPLE WEIGHT = 1/ PROBABILITY OF BEING CHOSEN] However, using weighting techniques does not solve the problem of multiplicity automatically as the different weights causes increased variance. (Kish, 1965, Section 11.7)

B. Weighting versus Unique Counting Rule

If the increased variance caused by weighting causes concern, an alternative method to deal with the problem of multiplicity is to employ the Unique Counting Rule (UCR). Using the UCR, each element is uniquely identified with one sampling unit. After choosing an individual to be included in the sample in the method outlined in the preliminary design [last section of paper], the interviewer will ask "Is this the first camp since January (or whatever time the survey was begun) that you have lived in?" If the MFW answers "yes", he is included in the survey. However, efficiency will decrease as the year progresses. It may be problematic at the end of the year to find households which have been unchanged since the beginning of the year. Thus, the UCR will (may) be more expensive than simply using weighting techniques. For a fixed budget, weighting would point to a large sample size because of precision losses due to variable weighting. It must also be noted that the UCR will tend to exclude the more mobile MFWs during the end of the survey period. This is an undesirable characteristic since for some survey measurements taken the more mobile MFWs may differ considerably from the less mobile MFWs. These MFWs may differ in age, sex, marital status, family size, etc. Thus, the decision must be made beforehand on whether cost effectiveness (using weighting) or decreased variance (using UCR) is desired.

C. How local studies dealed with multiplicity

Interestingly enough, local studies done on MFWs have not really addressed the issue of multiplicity. This is understandable since these studies traditionally take place over

relatively short periods of time and cover very limited areas. The problem of multiplicity, as compared to a nationwide study of a year's duration, is not as significant. For instance, the New York Migrant Health Interview Survey conducted during the summer of 1982 took place over an entire growing season. It was assumed that migrants tended to stay in the same residence when they worked in the survey area and thus double counting would not be a problem. However, since migrants are mobile, occupancy rates in the migrant housing units differed over time. To correct for this bias, a weighting procedure was used. (Chi, 1985).

V. MOBILITY

The severity of the "problem of multiplicity" is directly correlated with the migration patterns of MFWs. If MFWs are extremely mobile, multiplicity [recounting] becomes a significant issue. By understanding the migration patterns of MFWs [what motivates MFWs to move?, how often do they move?, where are they most heavily concentrated?] , we will hopefully be able to propose a design with the least amount of multiplicity and the highest rate of coverage.

A. Definition of migratory streams

Migratory streams, as illustrated on the next page, are the major patterns of migratory travel. Although MFWs do not strictly adhere to these "migratory streams", they provide a rough estimate of the migration process.

Definition of migratory streams.

East Coast Stream- From Florida and other states of the Southeast, workers start out in April, May, or in June. Generally they follow routes east of the Appalachians, although some go to the Midwest and even further west. They return to the Southeast in the fall when work "up North" is no longer available.

Mid-continent Stream- This stream starts out in all directions but with the majority going to Michigan, Ohio, Indiana, Illinois, Iowa, Minnesota and other midwestern states.

West Coast Stream- This stream originates in Texas and the Southwest. Some workers travel only within a single state, and others travel from Texas to the Pacific Northwest and back. (Johnston, 1985)

It is not surprising that the location of the MFW's permanent residence determines the entry point into the US. Jones (1984) found that California draws migrants from western Mexico while Texas draws migrants from NE Mexico. For example, Durango draws 3 times

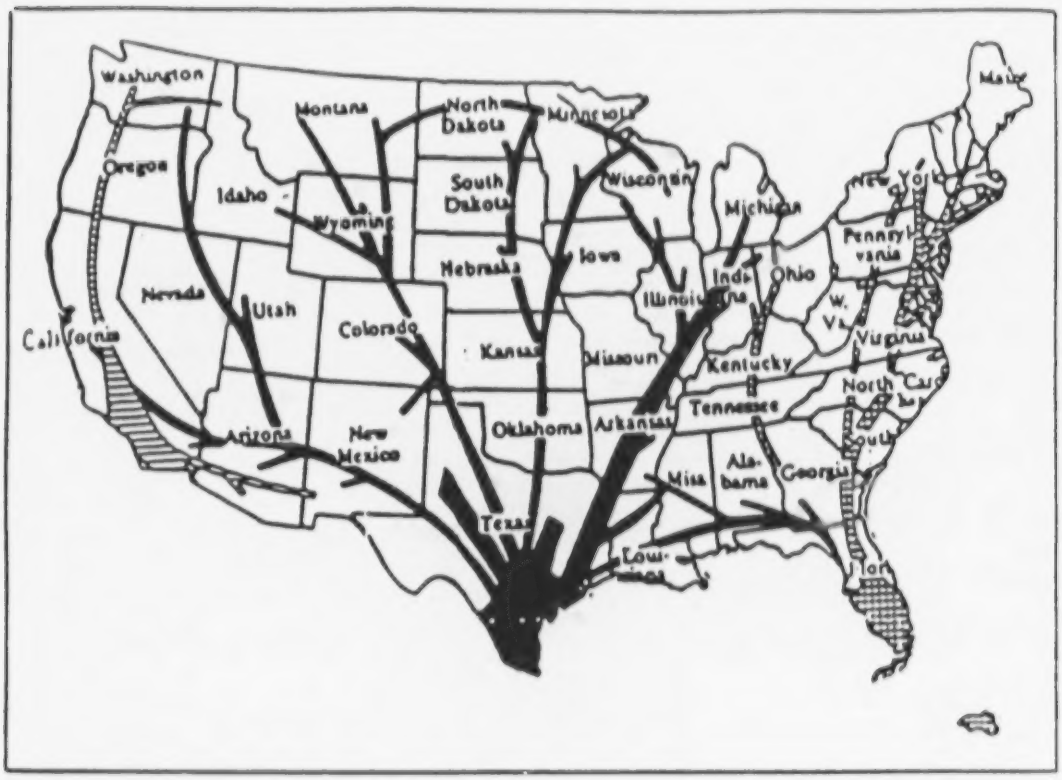


Chart 1

Travel patterns of seasonal migratory agricultural workers.

Source: Johnston, 1985

as many migrants to California as to Texas while San Luis Potosi draws 3 times as many migrants to Texas as to California. Convenience of location and information channels appear to be the causes of this high "channelization." [An information channel is the villagers' shared knowledge about the migrants from their regions. These Mexican villagers know who has migrated to the US, where they went, and their degrees of success. (Jones, 82).] Thus, it is possible that migrant camps will be populated by MFW's originating from the same general region.

The Jones and Massey datasets also indicate a certain degree of channelization. In the Massey dataset, 90.7% of the MFWs frequented only one US city when doing agricultual work. (see exhibit 17) In fact, four cities in California accounted for 55.7% of all the cities visited for the intent of doing agricultural work. Looking at between-year transitions (exhibit 14), one recognizes the sparsity of interstate movement. While 23.8% of these between-year movements are made within the same state, less than 1% of between-year movements are between different states.

The Jones dataset draws similar conclusions. In between-year transitions (exhibit 5), about 49% of MFWs are moving within the same US state. Only 9% of these MFWs are going to different states between seasons.

Thus, in both datasets , MFWs are moving to the same general area year after year. This is significant since agencies serving MFWs can anticipate (although not extremely accurately) the size of the MFW population in their localities. (The importance of obtaining Measures of Size will be stressed in the discussion of the proposed design.) Of course, MFWs are not always predictable and frequently do change their itineraries. Some of the motives for their changes in migration patterns are discussed below.

B. What motivates how MFWs move?

i. Agricultural Seasons

Migrant farmworkers will obviously pursue areas where there is a high demand for their labor. Thus, it stands to reason that peak months and duration of seasons are primary determinants of movement. Listed on the next page are the peak months and duration of seasons for the leading agricultural states.

ii. Weather

Harsh weather conditions, such as drought, hail, floods, etc. can force a MFW to change his itinerary. For example, last summer's drought (1988) caused decreased farm production. Since growers had a reduced demand, the MFWs traveled greater distances to find agricultural employment. In the Colorado Migrant Farmworker Health Survey, fruit-damaging frost and vegetable-damaging hail caused migrants to move to other areas or to other states. (Littlefield, Stout 1987) Thus, in the Colorado Study, occupancy rates in migrant camps were much lower than expected.

iii. Legal status

The study by Massey and Reichert (1979) compares migration routes for legal and illegal migrants from Guadalupe, Michoacan. It was suggested that legal migrants are more mobile than the migrants that are in the US illegally. This is of importance since estimates for the percentage of MFWs that are illegal are somewhat high ranging anywhere between 20-70% (Larson, 1982) [The implications of the new Amnesty Law on these figures will be discussed shortly.] Although their study is not limited to migrant farmworkers, the general results are still of interest.

Legal migration groups made an average of 1.2 moves in the US, compared to only 0.4 for illegal groups. The average number of moves for individual legal migrants is 1.8 while this figure if 0.3 for illegal migrants. The number of moves in this context was derived by

Table 1

Leading seasonal farm labor demand states, peak months of employment and duration of agricultural crop season.

State	Peak months	Duration of season
Louisiana	Jan-Feb	Jan-Dec
Florida	Mar-Apr	Jan-Dec
Illinois	May-Jun	May-Oct
Minnesota	May-Jun	May-Nov
Arkansas	May-Jun	May-Aug
New Mexico	May-Jun	Mar-Nov
Oklahoma	May-Jun	Jan-Dec
Nebraska	May-Jun	May-Aug
Colorado	May-Jun	Apr-Nov
North Dakota	May-Jun	May-Oct
Wyoming	May-Jun	May-Aug
Oregon	May-Jun	Jan-Dec
Connecticut	Jul-Aug	Apr-Nov
Massachusetts	Jul-Aug	May-Dec
New Jersey	Jul-Aug	May-Oct
Delaware	Jul-Aug	Apr-Oct
Maryland	Jul-Aug	May-Nov
Virginia	Jul-Aug	May-Nov
Alabama	Jul-Aug	May-Nov
North Carolina	Jul-Aug	Jan-Dec
South Carolina	Jul-Aug	Jun-Oct
Michigan	Jul-Aug	Apr-Nov
Wisconsin	Jul-Aug	May-Oct
Texas	Jul-Aug	Jan-Dec
Iowa	Jul-Aug	May-Nov
Kansas	Jul-Aug	Apr-Nov
Montana	Jul-Aug	Jan-Dec
Utah	Jul-Aug	May-Nov
Idaho	Jul-Aug	Mar-Nov
New York	Sep-Oct	May-Nov
Pennsylvania	Sep-Oct	Jan-Dec
Indiana	Sep-Oct	Apr-Oct
Ohio	Sep-Oct	May-Oct
Missouri	Sep-Oct	Jul-Oct
California	Sep-Oct	Jan-Dec
Washington	Sep-Oct	Jan-Dec
Arizona	Nov-Dec	Jan-Dec

Source: Public Health Service, Migrant Health Program, *1973 Migrant Health Program Target Population Estimates*, May 1975.

defining 13 geographic zones in the US. Six zones correspond to different areas within California. The rest represent particular areas within the states of Oregon, Washington, Florida, North Carolina, Indiana, Pennsylvania and Michigan. A move was defined as a movement between any 2 of these 13 zones. Two maps which illustrate the migration routes for legally migrating families and illegally migrating families are shown on the next page.

These patterns of movement illustrate that legal migrants have greater "spatial mobility" than illegal migrants. Illegals do not make any "trans- continental jumps." There is also a tendency for illegals to work in southern California or Florida and to remain there, without moving, while they are in the US. Also, illegal migrants tend to remain in the US longer than legal migrants. Illegal migrants are not as free to move back and fourth across the border since they fear detection. Below is a table listing the average duration of the last trip to the US by Migrant Parties from Guadalupe, Michoacan, 1978.

Legal families	9.4 months
Illegal families	12.4 months
Mixed families	12.6 months

Jones, Harris and Valdez (1984) found results very similar to Massey's and Reichert's. The average illegal migrant worked eleven months in the US followed by sixteen months in Mexico. However, these migrants are not limited to migrant farmworkers and 80% of the work trips were to Texas. (Most of Jones' studies focus on the south Texas region.)

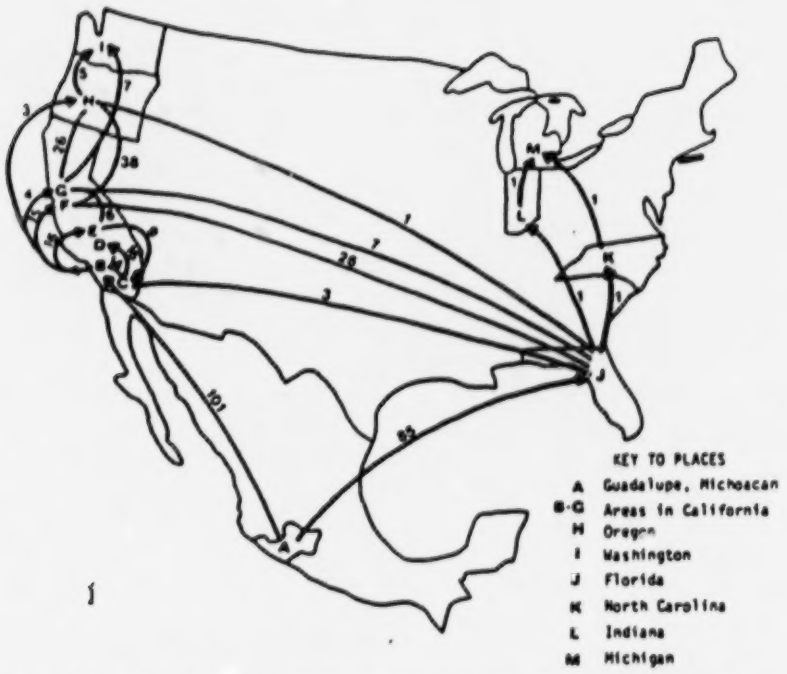


Figure 6.1 Migration routes for legally migrating family units from Guadalupe, Michoacán, to the United States (Specific place names have been deleted for reasons of confidentiality.)

Figure 6.2 Migration routes for illegally migrating family units from Guadalupe, Michoacán, to the United States

Source: "Patterns of Migration from a Mexican Town" by Massey,Reichert (1979)

VI. AN IN DEPTH DISCUSSION OF MIGRATION PATTERNS

In our analysis, we will be trying to answer the following questions: 1) How often do MFWs move? 2) What distance is travelled when MFWs seek new employment? 3) What states are most frequented? Migration patterns will help us to determine how close noncontiguous counties should be. In our proposed design, which will be discussed in great depth in section VIII of this paper, groups of noncontiguous counties will be the primary sampling units. If MFWs move relatively large distances, noncontiguous counties can be chosen closer together since multiplicity will not be a significant problem.

A. How often do MFWs move?

After investigating the three datasets (MSRTS, Massey, and Jones), we feel that MFWs are not as mobile as might be expected. (The literature on this topic is sparse since investigators [of MFWs] do not keep a detailed track of MFW movement.) The average lengths of time for agricultural jobs for the MSRTS, Massey data set, and Jones dataset are respectively 6.5 months, 9.5 months, and 5.4 months respectively. (See Exhibit 20) It is important to note that for the MSRTS, the lengths of time between movements were computed only for those students with a continuous record of school enrollments. Although one may speculate that those MFWs with discontinuous records may be more mobile, one must also note that "settled out" MFWs are also not included in this type of analysis. (Settled out MFWs may not move for years.)

When moves occurred within the year for the Massey dataset, it was primarily of an international nature. [Thus, these MFWs were not moving between agricultural jobs. In fact, in this dataset, only 1 MFW held more than one agricultural job in a year. As stated earlier,

this may be explained by our not having a complete within-year work history from the Massey data.] While only 7.5% of these moves are within the US, 73.1% of recorded moves are international. (See exhibits 10-11) Also, the fact that only 119 records out of over 700 illustrate within-year movement should stress the somewhat stationary nature of these MFWs in the Massey study.

In discussing the Jones dataset, only 39 out of 96 MFWs recorded holding more than one job in a year. However, the Jones MFWs are not as international as the Massey MFWs. (Only 37.8% of within-year movements involved Mexico and the US.) Nevertheless, the main point of interest is the movement between agricultural jobs since we are concerned with recounting MFWs who reside in different locations. Although 60.7% of within-year movements are in the US, only 19.7% of within-year movements were restricted to agriculturally related work. Hence, only 32.5% of within-year movements contained US borders are strictly agricultural. Thus, it appears that according to the Jones data, MFWs are not constantly moving between agricultural jobs. (See exhibits 4-6)

B. How far do MFWs travel?

Actual distances were calculated only for the MSRTS since the Massey and Jones data lacked the sufficient information necessary to estimate the mileage travelled between job sites. A considerable percentage of the movements in the Massey dataset was on an international basis while Jones' questionnaires did not list a majority of the cities traveled to by MFWs.

Generally speaking, the Migrant Student Record Transfer System (MSRTS) portrays MFWs as moving relatively large distances to reach their job destinations. Mean distances of 970.4 miles, 1694.6 miles, and 805.1 miles are traveled by the "continuous", "discontinuous" and "settled out" subsets respectively. "Continuous" and "discontinuous" refer to the status of a migrant student's school records while "settled out" refers to migrant students who are

no longer mobile. For more details on these different classifications, please review Appendix II. The formula used to calculate distance is:

SQUAREROOT[(TOXCEN-FROMXCEN)**2 + (TOYCEN-FROMYCEN)**2)] where

TOXCEN= center coordinates of longitude for county traveled to

FROMXCEN= center coordinates of longitude for county traveled from

TOYCEN= center coordinates of latitude for county traveled to

FROMYCEN= center coordinates of latitude for county travled from

For more details and the exact procedure for determining coordinates and distance traveled, refer to Appendix II.

SUMMARY OF DISTANCES TRAVELED (IN MILES) FOR MIGRANT CHILDREN

GROUP	N	MEAN DISTANCE TRAVELED (STD ERROR OF MEAN)	5 NUMBER SUMMARY		
MOBILE-CONTINUOUS RECORDS	1562	970.4 (23.81)	100%	MAX	3384.7
			75%	Q3	1576.6
			50%	MED	643.2
			25%	Q1	191.6
			0%	MIN	0
MOBILE-DISCONT. RECORDS	1767	1694.6 (23.91)	100%	MAX	3854.4
			75%	Q3	2649.6
			50%	MED	1584.4
			25%	Q1	965.6
			0%	MIN	0
"SETTLED OUT"	828	805.1 (31.34)	100%	MAX	3842.3
			75%	Q3	1355.7
			50%	MED	411.6
			25%	Q1	91.3
			0%	MIN	0

Although distances were not calculated for the Jones and Massey datasets, it is

possible to deduce certain conclusions. As stated previously, the Massey dataset was mostly international. Within a year, only 7.5% of all movements were made within the US. (See Exhibit 11.) Thus, the issue of how far these MFWs move to get to their next agricultural job seems irrelevant since these MFWs are for the most part nonmobile [when in the US].

The Jones dataset showed a considerable amount of interstate and intrastate movement. Within a year, 41.0% of MFWs moved within the same US state while 19.7% of MFWs moved between states. (See Exhibits 2-3). However, no MFWs held two agricultural jobs in the same state. Between-year transitions show very similar trends. While 49.2% of MFWs move within the same state, only 9.0% are moving between states. However, only 23.7% of MFWs are moving between agricultural jobs. (See Exhibits 4-6) Thus, in the Jones dataset, MFWs do not appear to be traveling short distances to labor at different agricultural jobs. Conclusion: Multiplicity does not appear to be as severe as originally expected. Thus, in the first stage of our proposed design, county groups do not need to encompass large areas. [See discussion of 4 stage design.] Furthermore, the plan to extend the survey period over a year's duration seems justified. [If mobility was a very serious problem, lengthening the time period would only compound the problem of double counting.]

C. Family vs solitary travelers

After a careful analysis of the MSRTS dataset, the two datasets from Robert Jones and Douglas Massey will be used to clarify these results. In the Jones and Massey datasets, special emphasis will be given to comparing the migration behaviors of families and solitary travelers. This is of particular importance since we must know the limitations of the MSRTS data, which tracks movement but only for those married MFWs with school age children. The Jones and Massey datasets will be answering the question "Can we infer the MSRTS results to the general migrant farmworker population?" Do families and solitary travelers

have similar migration patterns?"

Looking at Massey exhibits (# 15, 16, 18), the differences between MFWs traveling with and without children are not substantial. However , MFWs without children and working in an agricultural job seem to be slightly more mobile than MFWs traveling with children. (Note that MFWs traveling with children have a mean of 310.6 work days compared to a mean of 270.0 work days for those traveling without children.) The Jones dataset illustrates more differences between these two groups. At best, these particular data can only be thought of as preliminary since the sample sizes are small. (Demographic data was given for only a small portion of questionnaires.) Looking at exhibit 19, those with children seem more mobile than those traveling without children since their average length of work is of shorter duration. However, the mean workdays (150.2 for MFWs with children, 180.8 for MFWs without children) are not drastically different. The patterns of migration (see exhibits 7-8) for these two groups are somewhat similar but due to small sample sizes, it is difficult to generalize to the migrant farmworker population.

Unfortunately, the literature on this topic is sparse. In Massey and Reichert's (1979) study, legal migration groups made an average of 1.2 moves in the US compared to 1.8 moves for individual migrants. Illegal parties made an average of 0.4 moves while individual illegal migrants made 0.3 moves. Although, this shows differences between family travelers and solitary travelers, they are not considerable. One would expect, however, that solitary travelers would be even more mobile than families since they have less restraints. If this is the case, the lengths of time between movement for the MSRTS may be artificially high. It appears that the differences in mobility between the two groups of MFWs (families and solitary travelers) are minimal.

VII. DIFFICULTIES IN FINDING MIGRANT FARMWORKER HOUSING

A design study for a population of this type requires that several issues be resolved - one of which is the matter of actually "finding" the housing of MFWs. Since most migrant housing violates so many health standards, employers of MFWs may not allow health officials and/or house inspectors to know about the location of migrant camps. Furthermore, entry into a "private camp" may be denied to "outsiders." The property rights of camp owners supersede the individual's rights to have visitors, including visitors from health and social service agencies. (Johnston,1985) Thus, interviewers may have trouble gaining access to the migrant camps from the camp owners. Problems associated with locating migrant camps, training interviewers, selecting sampling units, language barriers, etc. will be discussed in conjunction with the proposed 4-stage design .

New Amnesty Law May Make Illegal Aliens More Elusive

The new Amnesty Law (defined below) may make MFWs who are in the US illegally even more secretive and wary of interviewers. Briefly, under this new law:

> Illegal immigrants in the US since before January 1, 1982 can seek amnesty. Furthermore, this law has a special provision for farmworkers. Immigrants who worked on US farms for at least 90 days from May 1, 1985 to May 1, 1986 can also seek amnesty. Employers who hire illegals act $10,000 fines. (INS, 1987)

Thus, those who cannot become legalized may be forced into "off-the-books businesses" where they could be more exploited. (Whitman, Hawkins 1987) Obviously, facing such heavy fines will make employers of illegal MFWs very hesitant in allowing interviewers access to their employees. [However, at the same time, MFWs who have become legalized may become more willing to be interviewed since they no longer have to

be clandestine.]

Thus, while the Amnesty law will allow some previously illegal MFWs amnesty, a new "underclass" may develop. As of May 1988, roughly half a million immigrants have applied for the program for seasonal agricultural workers. (Applebone, 1988) However, the actual number of legitimate MFWs applying for this program may be much lower since some Immigration and Naturalization Service officials believe this program is beset by fraud. For example, in Florida, officials think that more than half of the farmworkers applications contain fraudulent information. (Associated Press, 1988) Also, some legitimate MFWs may not have proof that they worked the required ninety days since their employers would not furnish them with letters or other documentation. Some MFWs could not afford the $185 application fee. (INS, 1987) This "underclass" cannot receive public assistance and cannot be hired legally. (Larson, 1982) The number of migrant farmworkers that will fall into this category is unknown.

Thus, actually finding migrant housing may be one of the most difficult aspects of implementing a national survey. However, as the discussion of the 4 stage proposed design will illustrate, the new Amnesty Law is only one obstacle in obtaining a complete listing of migrant housing units.

VIII. PROPOSED FOUR STAGE DESIGN

One method of implementing a national survey of migrant farmworkers is to apply a stratified, multi-stage cluster sample. The first stage utilizes area sampling methodology while the remaining stages utilize methods of list sampling. Along with the description of each stage, important issues related to selection in that stage will be discussed.

A. FIRST STAGE SAMPLING UNIT = COUNTY GROUPS

In the first stage, the continental US is divided into county groups. (There are around 3,000 counties.) Choosing the counties which will form the county group is not simply a function of geographic proximity. The density of MFWs and the migration patterns of MFWs are key indicators in selecting the composition of county groups.

i. Should counties with low concentrations of MFWs be included in the survey?

Obviously, MFWs are not evenly distributed throughout the US. Some areas are more densely populated than others. In fact, the Public Health Service Act, Title III, Part D has defined High Impact Areas as those which would legitimize the existence of a migrant health clinic.

> The term "high impact area" means a health service area or other area which has not less than four thousand migratory agricultural workers and seasonal agricultural workers residing within its boundaries for more than two months in any calendar year. In computing the number of workers residing in an area, there shall be included as workers the members of the families of such workers.

> The Secretary shall assign to high impact areas and other areas (where appropriate) priorities for the provision of assistance under this section to projects and programs in such areas. The highest priorties for such assistance shall be assigned to areas where the Secretary determines the greatest need exists. (Johnston, 1985)

Some states in the US are so sparsely populated with MFWs that they do not contain

any "High Impact Areas." The 357 migrant health clinics which exist are distributed within 35 states and Puerto Rico. (Schlegal, 1987) By excluding counties with low concentrations of MFWs, (the minimum population size obviously need not agree with the Public Health Service Act's) considerable time and money can be saved. The cost/benefit ratio involved in including any county group regardless of MFW density becomes drastically reduced as the MFW population size dwindles. Not only would it be very difficult to locate migrant camps in low density areas if indeed such camps existed, the expenses involved for including such a small population of MFWs cannot be justified. The only drawback to excluding counties with low concentrations of MFWs is the inability to achieve complete coverage. However, given the limited funds given to implement the survey, this may seem more appropriate.

ii. How to determine MFW population size to serve as measures of size for PSU selection

a. US Dept of Labor

On the county level, the USDL is the most reliable source of information regarding MFW population size. The USDL keeps a list of Agricultural Reporting Areas - county groups which group naturally into agricultural areas. A major source of the USDL's enumeration estimates is from local Rural Manpower Service Officials. However, the Rural Manpower Service depends on workers registering with the State Employment Commission Office or with the Annual Worker Plan. (Larson, 1982) This would underestimate the number of MFWs since those who entered the country illegally would fail to register.

The Rural Manpower Service has 261 agricultural reporting areas throughout the country. These areas are usually contiguous counties within state or single crop areas. On the 15th of each month, local RMS personnel maintain in season Farm Labor Reports [ES-233]. These contain estimates of seasonal hired employment in agriculture by crop activity

and origin of worker. (Schlenger, Ordrizek, Hallan, 1979)

These statistics are used by the US Department of Health and Human Services which maintains extensive listing of migrants and seasonal impact areas. This listing gives information on the location of health centers, estimated total number of migrants and seasonals in a county, agricultural areas, and agricultural area seasons. (US Dept. of Health and Human Services, 1987)

Unfortunately, not all states utilize the same enumeration methods. For instance, in some states, in lieu of using RMS data, the number of farms, crop acreage, or agricultural association records are used to determine base statistics. (Schlenger, Ordrizek, Hallan, 1979) Even with these inconsistencies, the data collected and maintained by the Department of Labor is the most complete source of information on MFWs that is available.

b. Census Bureau

The Census Bureau is another governmental agency which maintains statistics on population growth and trends. However, the Census Bureau has not made the special efforts necessary to get an adequate estimate of MFWs on the county level. In fact, the Census Bureau has five major disadvantages as a data source. These are 1) The 1980 Census made no distinctions between migrant seasonal farmworkers, permanent full-time workers, and permanent part-time farmworkers. 2) Minimal efforts were made in finding migrant camps. 3) During the 1980 Census period, many migrants were traveling to the Northwest and could not easily be reached by Census takers. 4) The Census listed the migrants' occupations during the week prior to April 1. Many migrants may not have started their work in agriculture as of this date and would not be categorized as a farmworker. 5) The Census does not indicate migrant travel patterns and temporary residences; only migrants' "homebases" are recorded. (Larson, 1982) This is the most significant disadvantage since the estimate of MFWs in their workplaces not their home residences is what is needed.

c. US Department of Agriculture

The USDA enumeration methodology is a probability sample via the Current Population Survey. From the 59,000 houses included in the CPS survey, only 1500 contained people who did hired farmwork. (Pollack, 1983) An even smaller number of these households (300) actually included MSFWs. (Larson, 1982) Thus, based on such small numbers, it would be impossible to provide county breakdowns of the data.

iii. Composition of the Primary Sampling Unit

After determining which counties qualify to be included in the sample, a decision must be made about the actual composition of the county groups. How large an area should the county groups cover? If MFWs are excessively mobile and are spatially restricted in their movements (i.e. move within county groups in the state), it may be preferable to have the PSU (county groups) expand over a vast area. By covering a larger area in a shorter time frame, the chance of double counting MFWs would be reduced. (The logic for the previous statement lies in the premise that in shorter time spans, MFWs would be less likely to migrate.) However, as the data from the MSRTS, Jones, and Massey indicate, MFWs do not seem to be as mobile as some of the literature on the MFW lifestyle would imply. The problem of multiplicity is not as significant as originally thought. Thus it seems feasible for the PSU to contain a smaller number of counties. The cost of field operations for sampling and interviewing would be reduced. Funds would be saved by the reduction in travel expenses, time in training interviewers, and administrative/operating expenditures.

B. SECOND STAGE SAMPLING UNIT=ENUMERATION DISTRICTS/BLOCK GROUPS

A stage for sampling between-county groups and migrant housing is necessary since no complete lists of migrant housing exists at the county level and they would be very

expensive to create. Thus, it would be too expensive to search for migrant housing directly from a county group since this predicates finding all migrant housing in a relatively vast area. By having the second stage sampling unit equal enumeration districts/block groups, we have a more confined area in which we must identify migrant housing.

As stated earlier, the Census Bureau does not distinguish migrant farmworkers from other agricultural workers. Thus, the Census Bureau would not have detailed information on the number of MFWs in an Enumeration District or Block Group. However, the Census Bureau can provide an approximation of the number of MFWs within an ED/BG through proxies. We can use the number of farmworkers in general and the presence of Hispanic surnames to indicate possible locations of MFWs. To select with PPS sampling, these proxy measures of size are necessary. Although we want a numerical value with a high degree of accuracy and using ED/BGs necessitates using approximations, the alternative method of listing all migrant camps in county groups is much too costly and cannot be justified.

Arrangements with the Census Bureau would be needed to obtain such data. It is probable that , because of the Bureau's other reporting priorities, retrieving the figures of the 1990 Census would be difficult very shortly after the Census was conducted. Thus, one could use the 1980 figures on the number of MFWs or wait for the 1990 figures and face possible delays.

C. THIRD STAGE SAMPLING UNIT= MIGRANT HOUSING CAMPS

The most difficult aspect of this stage is actually locating the migrant camps and other forms of migrant housing. Migrant camps are often hidden from view in obscure rural areas. Nevertheless, a list of migrant camps (and other habitations of MFWs), although incomplete, can be obtained by contacting the local public and private groups which supply assistance to MFWs.

Migrant health centers funded by the USPHS Office of Migrant Health would probably have the most complete source of information. In 35 states, the Migrant Health Program has more than 300 sites for health service delivery for migrant seasonal farmworkers. (Johnston, 1985). Among services that are used by migrants, medical facilities are among the most common. In fact, a pattern of increasing usage occurs the longer a migrant stays in the US. (Massey, 1984).

Utilizing other federal migrant and seasonal farmworker programs such as : Migrant Education, Migrant and Seasonal Farmworker Program, Migrant Headstart, and CETA may be of use. Migrant Education serves migrant children and operates the system for the transfer of migrant student records. The Migrant and Seasonal Farmworker Program helps MSFWs find alternative jobs in year round employment, improves working conditions, aids with self-help housing, and helps with education. The Administration for Children, Youth, and Familes (Migrant Headstart) provides health, educational, nutritional, etc. aid to impoverished migrant preschool children. The Comprehensive Employment Training Agency (CETA) helps economically disadvantaged people who are un- or underemployed by training and qualifying them for better jobs. (Catalog of Federal Domestic Assistance, 1988). However, in 1976, the US Department of Labor estimated that less than 10% of all MSFW families in the US received any public assistance- although the majority did qualify. (Jones, 1982).

Legal aid societies, church-affiliated relief organizations, migrant advocacy groups, local health departments, agricultural extension offices, grower associations, etc. are also plausible information sources. Information from several of these sources would have to be merged to produce the final frame.

Obviously, a complete listing of migrant housing units would be difficult to obtain even with the assistance of several agencies. Due to the sometimes unconventional nature

of MFW housing, agencies may not know that such dwellings even exist. For example, a MFW may rent a "shelter" which is no more than a leaky tent, a shack with cardboard walls, or even a converted school bus. (Allison, 1988)

Even the more traditional housing units are grossly substandard. Often overcrowded and filthy, lacking even running water, migrant housing is a source of many maladies. In a 1982 East Coast (Delaware, Maryland, and Virginia), the typical migrant housing unit was as follows. Living quarters were usually barrack-style buildings. Dwellings were overcrowded. One 12 foot by 12 foot room may be shared by six to eight people. A migrant camp ranged in size from 15 to 800 people. (Johnston,1985)

D. FOURTH STAGE SAMPLING UNIT= INDIVIDUAL MIGRANT "HOUSING UNIT"

The next task is to construct a list of the MFW housing units for each of the camps chosen in the previous stage. This part of the survey necessitates that an interviewer visits the camp sites. The quality of the interviewer will introduce bias since he will determine how accurate and complete the information will be. What now follows are step-by-step instructions to an interviewer on listing housing units.

1) Before going to the migrant camp, get a translator from a medical health center (or anyone familiar with the migrant camp) to come with you. To get higher response rates, the interviewer must gain the trust of the MFWs. It has been recommended that the interviewers should come from the same socioeconomic background as the MFWs. If possible, current or recent MFWs should be on the field staff. (Larson, 1982)

2) Walk around the camp and note the general housing. Inquire where the MFWs live. Do NOT include those who are not migrants- such as the crewleader.

3) Distinguish family units from those units containing single men. This is not necessarily an easy task since group quarters housing single men may be a dormitory or a house. Thus,

it is difficult to differentiate between 'family units' and 'single units' on the basis of appearance. It is wise to inquire about the residence status of the housing unit.

4) After determining where the migrants reside, one must "partition" the migrant camp (or any form of migrant settlement) into a series of housing units. The common definition of housing unit utilized by the Census Bureau is described below.

a. Housing unit- a housing unit is a group of rooms or a single room occupied or intended for occupancy as separate living quarters; that is,
   1. The occupants do not live and eat with any other persons in the structure
      AND
   2. There is either direct access from the outside or through a common hall
      OR
      complete kitchen facilities for the unit only.

   A housing unit may be occupied by a family group or an individual living alone. It may also be occupied by four or fewer unrelated persons.

b. Group quarters- Any single living unit in which five or more unrelated persons reside is defined as group quarters. This would apply to any structure or unit, including houses, apartments, barracks, and dormitories.

Unfortunately, the Census Bureau definition of a housing unit and group quarters is not always applicable in situations involving MFWs. The difficulty in defining a housing unit is finding an operational definition which is sufficiently general to apply to our sample. Generally speaking, housing units are permanent structures which uniquely identify a limited number of MFWs.

Although the Census definition for group quarters is adequate, it must be understood that a group quarters would not necessarily be our migrant housing unit. For example, if one hundred men live in a barracks, the barracks itself should not be defined as the housing unit. *The number of individuals per housing unit should not exceed some predetermined number. Instead, the barracks should be partitioned into a smaller units. If the barracks (or house, apartment, etc.) are divided into rooms [each of which houses MFWs], a room

would be defined as the housing unit. If there are no rooms, but each MFW has his own bed or mattress, groups of beds or mattresses could serve as a housing unit. The goal here is to identify some type of permanent structure which would exist in the same location at some future date (the date of the interviews). This is imperative since identifying housing units and conducting surveys will not be necessarily done on the same day. This is why housing units are selected -not individual MFWs- in this stage. Since migrants are mobile, there is always a possibility that they will not be at the same location at the time of the interview.

5.) When determining how group quarters should be partitioned into individual housing units, each housing unit should contain approximately the same number of people. This will serve to reduce the amount of variance and to insure consistency across camps.

6.) Label the "housing unit" with an identification number. An etching of the camp and the housing units which comprise the camp should be drawn. Distinctions should be made between family units and units of single men since these are our two substrata.

7.) Form a concatenated list of the migrant housing units in the two substrata (family units and units of single men). Housing units are then chosen by systematic sampling.

IX. CONCLUSION

This design study has identified and reviewed the primary problems associated with implementing a nationwide study of migrant farmworkers in the US. Issues such as defining a migrant farmworker, multiplicity, geographic mobility, occupational mobility, and elusiveness were examined. In the following paragraphs, we will summarize and highlight some of the key issues.

After reviewing several definitions of a migrant farmworker, it was ascertained that distinctions must be made between migrant seasonal farmworkers and migrant farmworkers. The terms "migrant" and "farmworker" also must be specifically defined since different connotations exist for these terms. Although this paper did not designate an official definition for a MFW, it did set one important limitation. A MFW must be employed in agriculture at the time of the actual survey to qualify as a MFW. This eliminates the need for a dual frame which would be very expensive and time-consuming.

The restriction placed on the definition of MFW is significant since preliminary findings indicate that MFWs are a dynamic population in terms of occupation. MFWs do not necessarily remain in agriculture on a continual basis; they may either labor in non-agricultural jobs or return to their homelands. Since individuals who are not MFWs in the beginning part of the year may be MFWs in the latter part (and vice versa), it was recommended that the study should extend over a twelve month period. Assuming the availability of resources, a year long study would have the distinct advantage of obtaining a higher coverage of the MFW population. It would also allow us to assess the health needs of MFWs in changing environmental conditions.

Since MFWs are also dynamic geographically, the problem of multiplicity was addressed. Mobility patterns from other studies and our three datasets (Migrant Student Record Transfer System, Jones, and Massey) were investigated to test the extent of multiplicity. This data showed that MFWs were not constantly moving between agricultural jobs. The average lengths of time for agricultural jobs for the MSRTS dataset, Massey dataset, and Jones dataset are 6.5 months, 9.5 months, and 5.4 months respectively. (To mitigate the effects of multiplicity, the Unique Counting Rule or Weighting are suggested.) Since multiplicity does not seem to be as prevalent as anticipated, the plan to extend the study to a year seems feasible. If multiplicity were a severe problem, it would have been more beneficial to shorten the length of the study since this reduces the chance of recounting.

In addition to the temporal analysis mentioned above, a spatial analysis of the MFW migration routes was completed. The MSRTS dataset indicated that MFWs are not moving short distances between jobs. (See Exhibit 21.) Thus, our county groups which comprise the Primary Sampling Unit could be close in proximity since the problem of MFWs moving between the county groups during the implementation of the survey [and thus being recounted] does not seem prominent. Although the distances traveled between jobs were not calculated for the Jones and Massey datasets, this data supports the findings of the MSRTS. These MFWs are not constantly moving short distances between agricultural jobs.

Finally, a proposed four-stage design is suggested. The sampling stages for the first, second, third, and fourth stages are county groups, enumeration districts/block groups, migrant camps, and migrant households respectively. The transitory nature of MFWs makes any proposed design difficult to implement. However, with cooperation from governmental agencies and organizations designed to aid migrant farmworkers, the

design suggested by William Kalsbeek, Ph.D. may be an adequate choice.

Although a nationwide study will not be a panacea for the occupational hazards associated with migrant farmwork, it may generate more concern for the plight of an often neglected and impoverished working group.

APPENDIX I

EXHIBITS 1-24 FOR THE JONES, MASSEY, AND MIGRANT STUDENT RECORD TRANSFER SYSTEM DATASETS

TABLE OF CONTENTS FOR EXHIBITS

EXHIBIT NUMBER	EXHIBIT DESCRIPTION
1	Jones, Within-year transitions, Employment Transitions
2	Jones, Within-year transitions, Geographical Transitions
3	Jones, Within-year transitions, General Geographical Transitions
4	Jones, Between-year transitions, Employment Transitions
5	Jones, Between-year transitions, Geographical transitions
6	Jones, Between-year transitions, General geographical transitions
7	Jones, Between-year transitions, Employment transitions, Children traveling
8	Jones, Between-year transitions, Employment transitions, No children traveling
9	Massey,Within-year transitions, Employment Transitions
10	Massey, Within-year transitions, Geographical Transitions
11	Massey, Within-year transitions, General Geographical Transitions
12	Massey, Between-year transitions, Employment Transitions
13	Massey, Between-year transitions, Geographical transitions
14	Massey, Between-year transitions, General geographical transitions
15	Massey, Between-year transitions, Employment transitions, Children traveling
16	Massey, Between-year transitions, Employment transitions, No children traveling
17	Massey, Number of different cities traveled to for agricultural work
18	Massey, Mean number of days worked for those with and without children
19	Jones, Mean number of days worked for those with and without children
20	All datasets, Mean number of days worked
21	MSRTS- Summary of distances traveled
22	MSRTS- Graph of lengths of time between movement
23	MSRTS- States migrants chose to "settle out"
24	MSRTS- Reasons for gaps in discontinuous data

FIRST JOB OF YEAR	LAST JOB OF YEAR: Non-agric	Agric	
Non-Agriculture	18 (29.5%)	17 (27.9%)	35 (57.4%)
Agriculture	14 (22.9%)	12 (19.7%)	26 (42.6%)
	32 (52.4%)	29 (47.6%)	61

Jones data
Within-year transitions
Sample size is 39

EXHIBIT 1

LAST JOB OF YEAR

FIRST JOB OF YEAR	TEXAS	REST OF THE US	MEXICO	
TEXAS	22 (36.1%)	6 (9.8%)	8 (13.1%)	36 (59.0%)
REST OF THE US	4 (6.6%)	5 (8.2%)	1 (1.6%)	10 (16.4%)
MEXICO	14 (23.0%)	0 (0.0%)	1 (1.6%)	15 (24.6%)
	40 (65.7%)	11 (18.0%)	10 (16.3%)	61

Jones data
Within-year transitions
Includes agricultural and non-agricultural jobs
Sample size is 39

EXHIBIT 2

TYPE OF MOVEMENT	FREQUENCY	PERCENT
US STATE TO DIFFERENT US STATE	12	19.7%
US STATE TO SAME US STATE	25	41.0%
US STATE TO MEXICAN STATE	9	14.8%
MEXICO TO US STATE	14	23.0%
MEXICAN STATE STATE TO MEXICAN STATE	1	1.6%
	61	100.0%

Jones data
Includes agricultural and non-agricultural jobs
Within year transitions
Sample size is 39

EXHIBIT 3

LAST JOB OF PREVIOUS YEAR	FIRST JOB OF SUCCEEDING YEAR: Non-agric	Agric	
Non-Agriculture	39 (22.0%)	51 (28.8%)	90 (50.8%)
Agriculture	45 (25.4%)	42 (23.7%)	87 (49.1%)
	84 (47.4%)	93 (52.5%)	177

Jones data
Between-year transitions
Sample size is 81

EXHIBIT 4

LAST JOB OF PREVIOUS YEAR	FIRST JOB OF SUCCEEDING YEAR: TEXAS	REST OF THE US	MEXICO	
TEXAS	77 (43.5%)	8 (4.5%)	25 (14.1%)	110 (62.1%)
REST OF THE US	5 (2.8%)	13 (7.3%)	5 (2.8%)	23 (12.9%)
MEXICO	34 (19.2%)	3 (1.7%)	7 (4.0%)	44 (24.9%)
	116 (65.5%)	24 (13.5%)	37 (20.9%)	177

Jones data
Between-year transitions
Includes agricultural and non-agricultural jobs
Sample size is 81

EXHIBIT 5

TYPE OF MOVEMENT	FREQUENCY	PERCENT
US STATE TO DIFFERENT US STATE	16	9.0%
US STATE TO SAME US STATE	87	49.2%
US STATE TO MEXICAN STATE	30	16.9%
MEXICO TO US STATE	37	20.9%
MEXICAN STATE STATE TO MEXICAN STATE	7	4.0%
	177	100.0%

Jones data
Between-year transitions
Includes agricultural and non-agricultural jobs
Move is last job of previous year to first job of next year
Sample size is 81

EXHIBIT 6

LAST JOB OF PREVIOUS YEAR	FIRST JOB OF SUCCEEDING YEAR: Non-agric	Agric	
Non-Agriculture	9 (22.5%)	11 (27.5%)	20 (50.0%)
Agriculture	9 (22.5%)	11 (27.5%)	20 (50.0%)
	18 (45.0%)	22 (55.0%)	40

Jones data
Between year transitions
Sample size is 19
These MFWs have children accompanying them

EXHIBIT 7

LAST JOB OF PREVIOUS YEAR	FIRST JOB OF SUCCEEDING YEAR: Non-agric	Agric	
Non-Agriculture	8 (22.9%)	14 (40.0%)	22 (62.9%)
Agriculture	5 (14.3%)	8 (22.9%)	13 (37.2%)
	13 (37.2%)	22 (62.9%)	35

Jones data
Between year transitions
Sample size is 22
These MFWs have no children accompanying them

EXHIBIT 8

FIRST JOB OF YEAR	LAST JOB OF YEAR: Non-agric	LAST JOB OF YEAR: Agric	
Non-Agriculture	27 (22.7%)	28 (23.5%)	55 (46.2%)
Agriculture	63 (52.9%)	1 (0.8%)	64 (53.7%)
	90 (75.6%)	29 (24.3%)	119

Massey data
Within-year transitions
Sample size is 47

EXHIBIT 9

FIRST JOB OF YEAR	LAST JOB OF YEAR: CALIFORNIA	REST OF THE US	MEXICO	
CALIFORNIA	5 (4.2%)	0 (0%)	52 (43.7%)	57 (47.9%)
REST OF THE US	0 (0%)	4 (3.4%)	7 (5.9%)	11 (9.3%)
MEXICO	28 (23.5%)	0 (0%)	23 (19.3%)	51 (42.8%)
	33 (27.7%)	4 (3.4%)	82 (68.9%)	119

Massey data
Within-year transitions
Includes agricultural and non-agricultural jobs
Sample size is 47

EXHIBIT 10

TYPE OF MOVEMENT	FREQUENCY	PERCENT
US STATE TO DIFFERENT US STATE	3	2.5%
US STATE TO SAME US STATE	6	5.0%
US STATE TO MEXICAN STATE	59	49.6%
MEXICO TO US STATE	28	23.5%
MEXICAN STATE STATE TO MEXICAN STATE	23	19.3%
	119	100.0%

Massey data
Includes agricultural and non-agricultural jobs
Within-year transitions
Sample size is 47

EXHIBIT 11

LAST JOB OF PREVIOUS YEAR	FIRST JOB OF SUCCEEDING YEAR: Non-agric	Agric	
Non-Agriculture	358 (55.6%)	94 (14.6%)	452 (70.2%)
Agriculture	72 (11.2%)	120 (18.6%)	192 (29.8%)
	430 (66.8%)	214 (33.2%)	644

Massey data
Between-year transitions
Sample size is 109

EXHIBIT 12

FIRST JOB OF SUCCEEDING YEAR

LAST JOB OF PREVIOUS YEAR	CALIFORNIA	REST OF THE US	MEXICO	
CALIFORNIA	144 (22.4%)	4 (.6%)	68 (10.6%)	216 (33.6%)
REST OF THE US	1 (.2%)	10 (1.6%)	6 (0.9%)	17 (2.7%)
MEXICO	88 (13.7%)	8 (1.2%)	315 (48.9%)	411 (63.8%)
	233 (36.3%)	22 (3.4%)	389 (60.4%)	644

Massey data
Between-year transitions
Includes agricultural and non-agricultural jobs
Move is last job of previous year to first job of next year
Sample size is 109

EXHIBIT 13

TYPE OF MOVEMENT	FREQUENCY	PERCENT
US STATE TO DIFFERENT US STATE	6	0.9%
US STATE TO SAME US STATE	153	23.8%
US STATE TO MEXICAN STATE	74	11.5%
MEXICO TO US STATE	96	14.9%
MEXICAN STATE STATE TO MEXICAN STATE	315	48.9%
	644	100.0%

Massey data
Includes agricultural and non-agricultural jobs
Between-year transitions
Move is last job of previous year to first job of next year
Sample size is 109

EXHIBIT 14

FIRST JOB OF SUCCEEDING YEAR

LAST JOB OF PREVIOUS YEAR	Non-agric	Agric	
Non-Agriculture	101 (49.0%)	43 (20.9%)	144 (69.9%)
Agriculture	11 (5.3%)	51 (24.8%)	62 (30.1%)
	112 (54.3%)	94 (45.7%)	206

Massey data
Between-year transitions
Sample size is 43
These MFWs have children accompanying them

EXHIBIT 15

LAST JOB OF PREVIOUS YEAR	FIRST JOB OF SUCCEEDING YEAR: Non-agric	Agric	
Non-Agriculture	257 (58.7%)	51 (11.6%)	308 (70.3%)
Agriculture	61 (13.9%)	69 (15.8%)	130 (29.7%)
	318 (72.6%)	120 (27.4%)	438

Massey data
Between year transitions
Sample size is 81
These MFWs have no children accompanying them.

EXHIBIT 16

NUMBER OF US TRIPS

NUMBER OF US CITIES	1	2	3	4	5	6	7	8	
1	50	17	8	8	5	2	7	1	98
2		3	2	1			1		7
3					1	1			2
4									0
5						1			1
	50	20	10	9	6	4	8	1	108

Source: Massey's Person-year Data in Mexico
108 represents the number of MFWs in Massey's dataset
1 missing

EXHIBIT 17

MASSEY DATA

WITHIN-YEAR MOVEMENTS

STATUS	N	MEAN (IN DAYS)	STD DEV OF MEAN
Children	49	168.4	12.82
No children	70	181.7	10.12

BETWEEN-YEAR MOVEMENTS

STATUS	N	MEAN (IN DAYS)	STD DEV OF MEAN
Children	206	315.7	6.34
No children	438	312.6	4.47
All	644	313.6	3.72

ALL MOVEMENTS

STATUS	N	MEAN (IN DAYS)	STD DEV OF MEAN
Children & agricultural job	99	310.6	8.33
No children & agricultural job	177	270.0	8.21

The mean number of days represents the mean length of employment. "Children" and "No children" refer to the travel status of the migrant. These terms indicate if the MFW is traveling with or without children.

EXHIBIT 18

JONES DATA

BETWEEN-YEAR MOVEMENTS

STATUS	N	MEAN (IN DAYS)	STD DEV OF MEAN
Children	40	150.2	16.99
No children	35	180.8	18.52

Since demographic data is sparse in the Jones dataset, statistics for within-year movements are not given. Small sample sizes prevented further breakdowns into agricultural and non-agricultural jobs.

EXHIBIT 19

DESCRIPTION	N	MEAN (IN DAYS)	STD DEV OF MEAN
MIGRANT STUDENT RECORD TRANSFER SYSTEM	645	195.4	5.39
MASSEY & AGRICULTURAL	276	284.6	4.89
MASSEY & NONAGRICULTURAL	596	300.5	4.31
JONES & AGRICULTURAL	238	161.3	7.54
JONES & NON-AGRICULTURAL	116	150.2	9.82

The mean number of days represents the mean length of employment for the Massey and Jones datasets. For the MSRTS, the mean number of days represents the length of time between movements.

The mean number of days for the MSRTS was calculated only for migrant students with continuous school records.

EXHIBIT 20

SUMMARY OF DISTANCES TRAVELED (IN MILES) FOR MIGRANT CHILDREN

GROUP	N	MEAN DISTANCE TRAVELED (STD ERROR OF MEAN)	5 NUMBER SUMMARY		
MOBILE-CONTINUOUS RECORDS	1562	970.4 (23.81)	100%	MAX	3384.7
			75%	Q3	1576.6
			50%	MED	643.2
			25%	Q1	191.6
			0%	MIN	0
MOBILE-DISCONT. RECORDS	1767	1694.6 (23.91)	100%	MAX	3854.4
			75%	Q3	2649.6
			50%	MED	1584.4
			25%	Q1	965.6
			0%	MIN	0
"SETTLED OUT"	828	805.1 (31.34)	100%	MAX	3842.3
			75%	Q3	1355.7
			50%	MED	411.6
			25%	Q1	91.3
			0%	MIN	0

DISTANCE (miles)	MOBILE-CONTINUOUS (N=1560)	MOBILE-DISCONTINUOUS (N=1767)	"SETTLED OUT" (N=828)
0	6.0	2.0	10.7
1-49	2.8	0.7	5.2
50-249	20.0	5.0	25.5
250-499	16.7	8.7	11.9
500-749	7.8	3.5	8.5
750-999	11.0	4.5	7.0
1000+	35.7	75.6	31.2
	100%	100%	100%

EXHIBIT 21

BEST COPY AVAILABLE

EXHIBIT 22

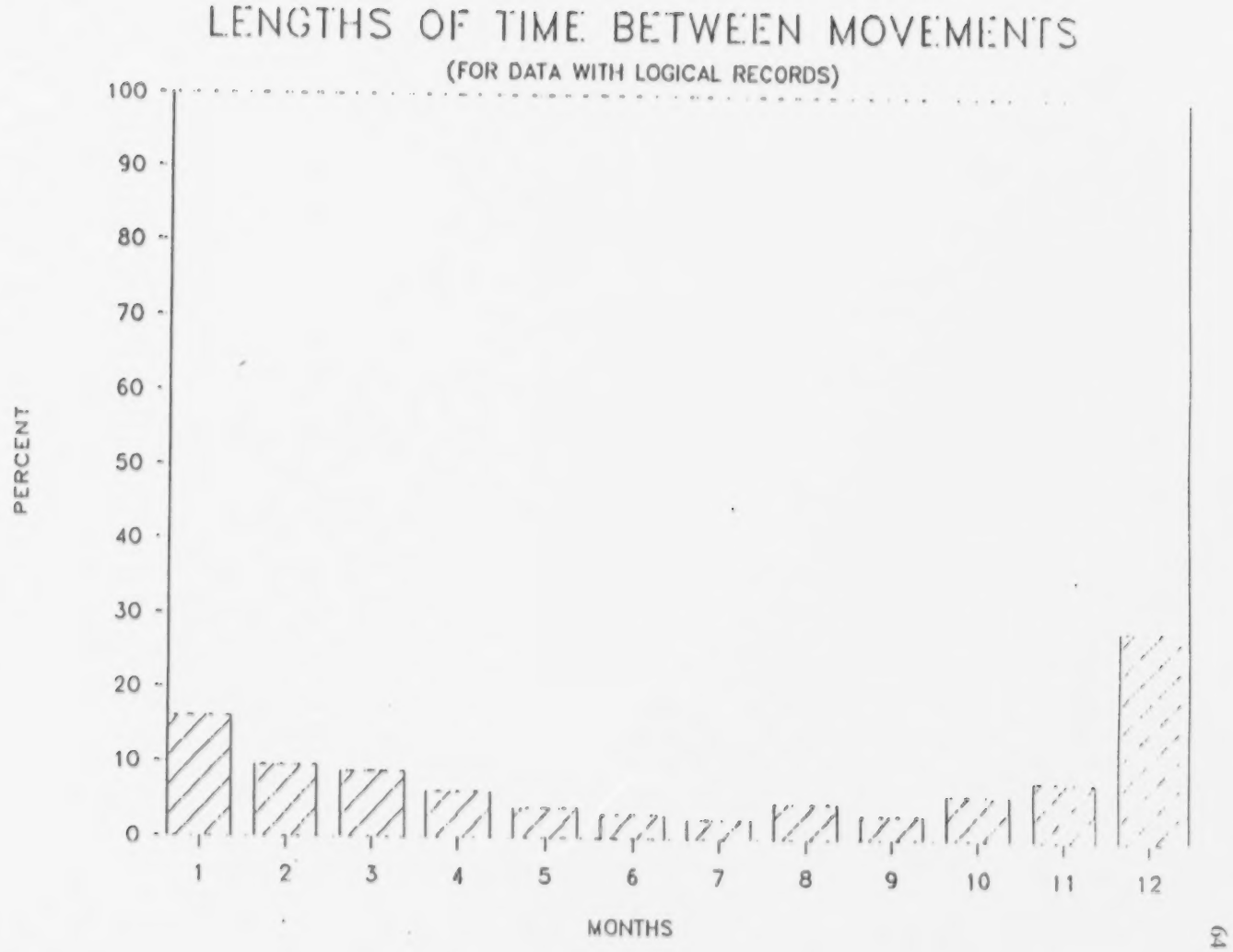


EXHIBIT 23

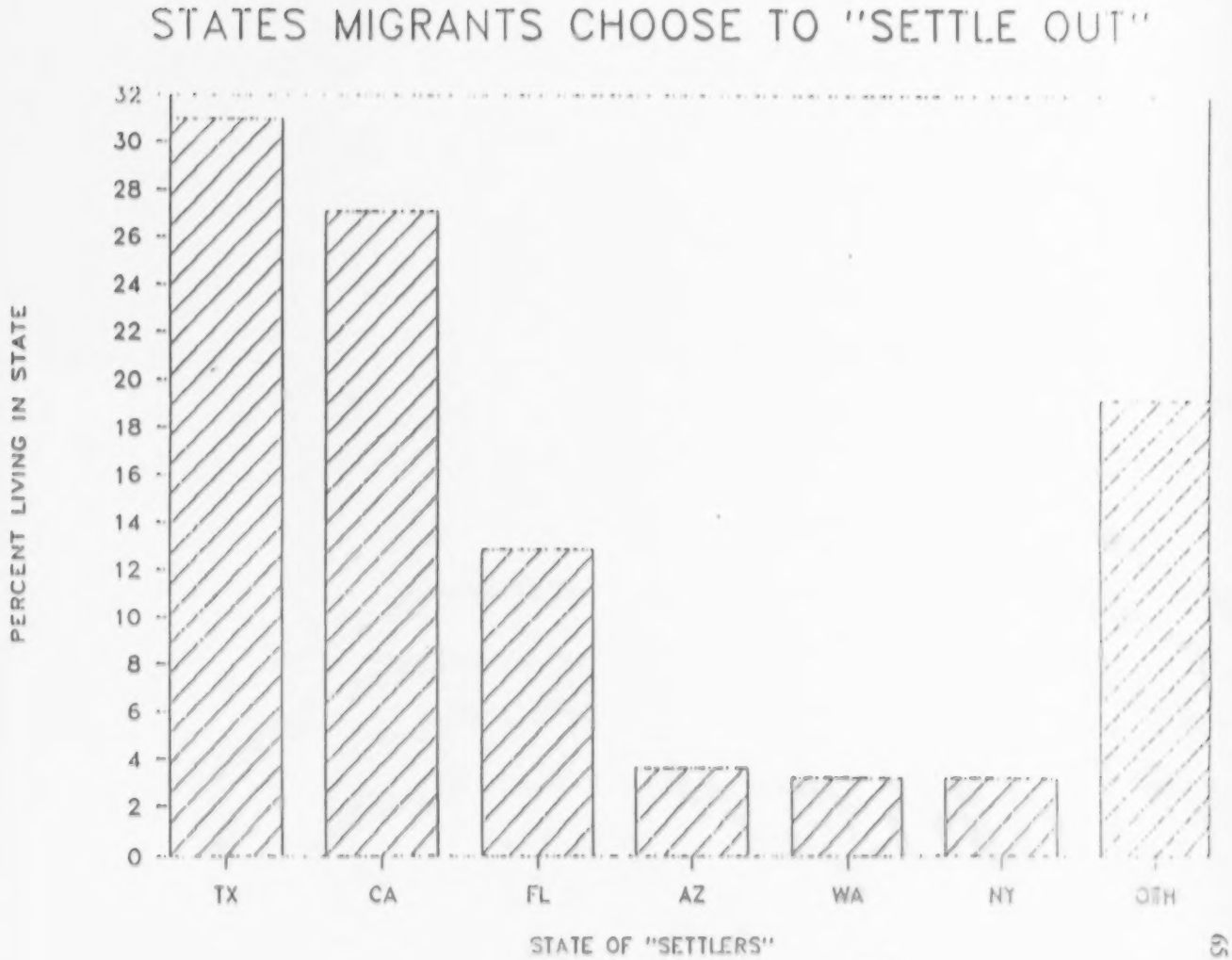
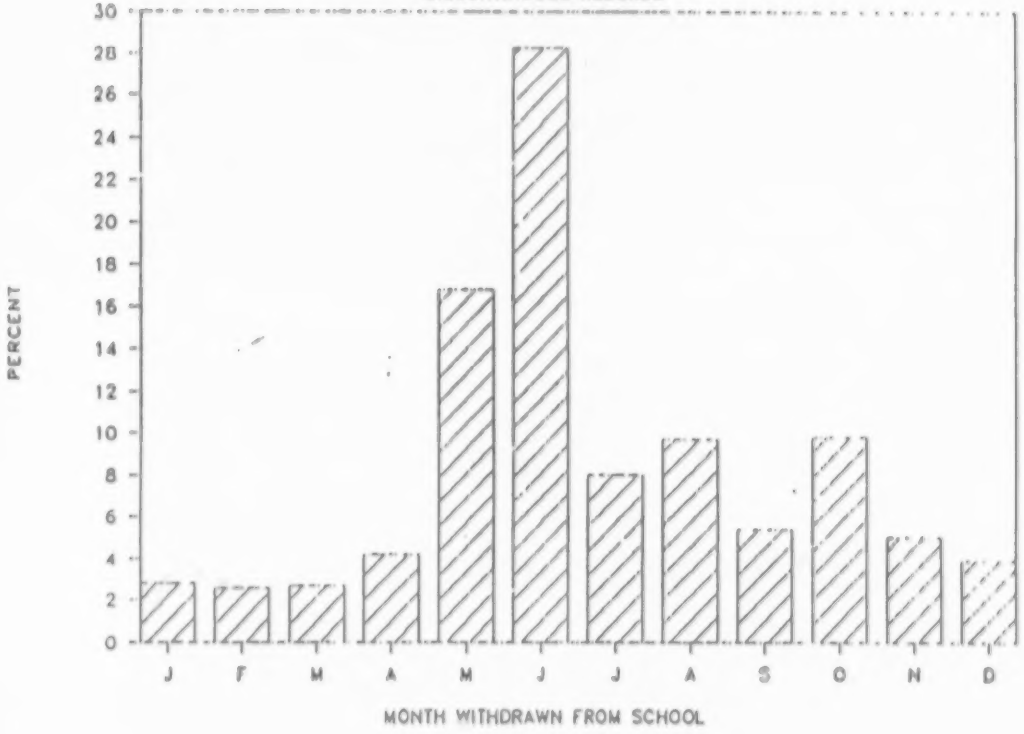

STATES MIGRANTS CHOOSE TO "SETTLE OUT"
PERCENT LIVING IN STATE
32
30
28
26
24
22
20
18
16
14
12
10
8
6
4
2
0
TX
CA
FL
AZ
WA
NY
OTH
STATE OF "SETTLERS"

EXHIBIT 24

EXPLANATION FOR SOME OF THE DISCONTINUOUS RECORDS



Many migrant children withdrew from school during the summer months. This caused many of the records to be discontinuous.

APPENDIX II

DETAILS ON THE MIGRANT STUDENT RECORD TRANSFER SYSTEM DATASET HOW THE DATASET WAS PARTITIONED AND HOW DISTANCE WAS CALCULATED

THE MIGRANT STUDENT RECORD TRANSFER SYSTEM DATASET (MSRTS)

As stated earlier, a systematic list sampling technique was employed in choosing a random sample of migrant children from the MSRTS. The tenth child listed in the database was chosen and every 50th child thereafter. This technique resulted in a sample size of 2500 migrant children or roughly 2% of the migrant population. Our sample size was further reduced when certain conditions were not met. If the parents of the migrant child were not engaged exclusively in agricultural work or if the child's records were from the wrong time frame, the migrant child was not included in analysis. Peculiarities in data, such as identical migration dates with different destinations for each date were also noted and excluded from study. Also, several records were deleted because they contained repetitive information. A migrant child may have multiple records which were identical except for the enrollment data in school. [Dates of movement were identical so no migration was taking place.]

Partitioning the dataset

The biggest drawback of the MSRTS data system is the lack of information on summer migrants. During the summer, migrant students do not attend school so the Migrant Education Programs are not in operation. (Larson,1982). Thus, many of the records seem to have missing gaps of data. To explain further, there are multiple records for each migrant child. Each record contains the city and state a migrant child moved from as well as the city and state the migrant child moved to. However, as previously noted, some of the records are noncontinuous. For instance, one record may have Houston, Texas as the last location

moved to. The next record of the same migrant child may have Hope, Arkansas as the location he is moving from. Thus, the sequential records are not continuous in many cases. (2246 records with a sample size of 598) fall into this category. (See Exhibit 24) Records which contain a logical sequence of locations comprises the second group. (1683 records with a sample size of 620) fall in this category. The third and final group is comprised of migrants who have "settled out". These MFWs are currently not migrating even though they may have done so in the past. (See Exhibit 23) 1053 records with a sample size of 1053* fall into this category. *Only one record per individual was needed for analysis since no migration was taking place. Multiple records for migrant children in this category contained repetitive information. A group of records for a migrant child differed only in enrollment and withdrawl dates from school.

Distance Traveled

How far do MFWs travel to reach a new destination? Determining the mileage traveled for thousands of records proved to be a difficult task since the only information available was the city and state "moved from", the city and state "moved to", and half of the zipcodes. To calculate distance, we had to determine the county of residence based on the city and state and translate this into longitude-latitude coordinates. The only dataset readily available with coordinates is the data set entitled Counties provided by SAS Institute, Cary. NC with the SASGRAPH package. SASGRAPH also has a dataset, US Cities, but this is a very small dataset containing a very limited number of locations. Thus, any distance measures made in our analysis are taken from the center of each county. This proved to be a good estimate when several of the distances were checked by hand.

Merging Datasets

Since the MSRTS dataset did not contain the numeric county and state codes available in Counties, the two datasets could not be merged directly. A third dataset .

Syssas.Zip.County served as an intermediary between the two datasets. This intermediary contained the variable Zipcode which would enable it to be merged with the MSRTS dataset. Syssas.Zip.County also contained the variables for county code and state code common to the Counties dataset. The merger between the MSRTS dataset and Syssas.Zip.County created a new dataset that could now be merged with the 'Counties' dataset by county and state. Unfortunately, the MSRTS dataset only supplied zipcodes for half of its cities. However, to calculate distance, the zipcodes for all the cities must be known. (This is only logical since by knowing the zipcode, coordinates can be calculated.) Using books supplied by the US Post Office, the zipcodes for hundreds of cities were manually looked up and then entered into the computer. However, not all cities were listed in the post office book. This was probably due to errors in data recording by the MSRTS. Zip codes for foreign countries were not determined since the Counties dataset is limited to the continental US. Thus, distances between the US and foreign countries were not calculated. The MSRTS dataset was merged indirectly with the Counties dataset using the zipcodes to provide a common key. Thus, the coordinates for the "from destination" and the coordinates for the "to destination" were known. How the distance was calculated given this information is illustrated on the next page.

Formulas for Determining Distance

The SAS data set, Counties, contains the unprojected coordinates of the county boundaries for all 48 continental states in the U.S. along with Hawaii and Puerto Rico. The formula for distance is:

SQUAREROOT((TOXCEN-FROMXCEN)**2 +(TOYCEN-FROMYCEN)**2))

TOXCEN= Center of county traveling to in degrees longitude
FROMXCEN= Center of county traveling from in degrees longitude
TOYCEN= Center of county traveling to in degrees latitude
FROMYCEN= Center of county traveling to in degrees latitude

COUNTIES supplies the following information:
X= Unprojected longitude in radians
Y= Unprojected latitude in radians

STEPS INVOLVED FOR GETTING DISTANCE: (Muehrck, 1978)

1) Convert radians into degrees. Multiply radians by 57.3 to convert into degrees.

2) To get the coordinates for the center of each county, the minimum and maximum longitude and the minimum and maximum latitude had to be determined.

Minimum longitude in degrees = (57.3) x MIN(X)= MNX

Maximum longitude in degrees = (57.3) x MAX(X)= MXX

Minimum latitude in degrees = (57.3) x MIN(Y)= MNY

Maximum latitude in degrees = (57.3) x MAX(Y)= MXY

3) Since longitude changes as the cosine of latitude,

one degree of longitude = cos(latitude) x 69.172.

Thus,

MNX1 = MNX x [cos(MNY) x 69.172]

MXX1 = MXX x [cos(MXY) x 69.172]

4) Formula to get the center coordinates of the county.

XCEN = ((MXX1-MNX1/2) + MNX1)

YCEN = ((MXY-MNY/2) + MNY)

5) FROMXCEN = Center coordinates of longitude for county traveling from

FROMYCEN = Center coordinates of latitude for county traveling from

TOXCEN = Center coordinates of longitude for county traveling to

TOYCEN = Center coordinates of latitude for county traveling to

6) DISTANCE FORMULA =

SQUAREROOT [(TOXCEN-FROMXCEN)**2 + (TOYCEN-FROMYCEN)**2)

BIBLIOGRAPHY

Allison, Sue. "Fruit Tramps". Life, December 1988, pp.127-132.

Applebone, Peter. "Built on Compromise, Amnesty Program Ends in Mixed Review." Raleigh News and Observer, May 8, 1988.

Associated Press. "Farmworkers Spill the Beans on Amnesty Scam." USA Today, May 8 1988.

Chi, Peter. "A Note on Sampling Migrant Farmworkers". The Rural Sociologist, May 1985, pp. 158-162.

Immigration and Naturalization Service. "What You Need for Legalization". USA Today, November 24, 1987.

Johnston, Helen. Health for the Nation's Harvesters. National Migrant Worker Council, Inc., 1985.

Jones, Richard. "Changing Patterns of Undocumented Mexican Migrants to South Texas". Social Science Quarterly, 65(2), June 1984, pp. 465-481.

Jones, Richard. "Channelization of Undocumented Mexican Migrants to the US". Economic Geography, vol 58, 1982, pp.156-176.

Jones, Richard. "Using US Immigration Data: Undocumented Migration from Mexico to South Texas". Journal of Geography, 83(2), 1983, pp. 58-64.

Jones, Richard, Harris, R.J., and Valdez, A. "Occupational and Spatial Mobility of Undocumented Migrants from Dolores Hidalgo, Guanajuato". Patterns of Undocumented Migration: Mexico and the US, (Totowa, NJ: Rowman and Allanheld, 1984).

Jones, Richard and Murray, William. "Occupational and Spatial Mobility of Temporary Mexican Migrants to the U.S.: A Comparative Analysis." International Migration Review, Volume XX, Number 4, Winter 1986, pp. 973-985.

Kish, Leslie. Survey Sampling, John Wiley & Sons, Inc. 1965.

Larson, Alice. A Guidebook to Gathering Data on Migrant and Seasonal Farmworkers, University of Washington, 1982.

Littlefield, Carla and Stout, Charles. Colorado Migrant Farmworker Health Survey Preliminary Report, Colorado Migrant Health Program, 1987.

Littlefield, Carla and Stout, Charles. "A Survey of Colorado's Migrant Farmworkers Access to Health Care". Migrant Health Clinical Supplement, February/March, 1988.

Massey, Douglas. "Patterns of US Migration from a Mexican Sending Community. A Comparison of Legal and Illegal Migrants. International Migration Review, 13 (1979): 599-623.

Massey, Douglas. "Settlement Process Among Mexican Migrants to the US: New Methods and Findings". Working Paper, Population Studies Center, University of Pennsylvania, 1984.

Massey, Douglas. "Understanding Mexican Migration to the US". American Journal of Sociology, Volume 92 Number 6 ,May 1987, pp.1372-1403.

Massey, Douglas and Reichert, Joshua. "Patterns of US Migration from a Mexican Town". Patterns of Undocumented Migration and the United States, Edited by Richard Jones, Rowman and Allanheld publishers, 1979.

McCoy, Terry. "Impact of US Temporary Worker Programs on Caribbean Development: Evidence from H-2 Workers in Florida Sugar", Migration and Development in the Caribbean- The Unexpected Correction , Edited by Robert A. Pastor, 1985, pp. 178-206.

Muehrck, Phillip. Map Use Reading , Analysis and Interpretation, Madison, Wisconsin, Jp Publications, 1978.

Okador, Y. , Richards, Slesinger. Migrant Worker Studies: A Critical Review of Methodologies, Center for Demography and Ecology, University of Wisconsin-Madison, June 1982.

Pollack, Susan. "The Hired Farm Working Force of 1983: A Statistical Profile", Agricultural Report No. 554, Economic Research Service USDA, 1983.

Quillan, Martha. "NC Must Improve Health Conditions of Migrant Workers". Raleigh News and Observer, May 16, 1988.

Schlegal, Helen. "National Migrant Referral Project: Facilitating Migrant Health Care for 15 Years". Migration Today, Volume XII, No. 415, 1987.

Schlenger, Ordrizek, Hallan. "An Examination of Methods Used to Estimate the Number of Migrant Farmworkers." The Farmworker Journal, Winter 78-79, pp.23-32.

US Dept of Health and Human Services. Migrant and Seasonal Impact Areas, National Clearing House for Primary Care Information, 1987.

APPENDIX B

Provisional Design for a Natural Probability Sample of Migrant Farm Workers

PROVISIONAL DESIGN FOR A NATIONAL PROBABILITY SAMPLE OF MIGRANT FARM WORKERS

A national survey of migrant farm workers (MFWs) as currently envisioned would be applied to a stratified, multi-stage cluster sample. Selecting this sample would utilize standard methods of area sampling through the first stage of selection and then of list sampling in the remaining stages.

General issues:
(1) Which of several definitions of "migrant farm worker" adopted by the various federal agencies and groups would be used for operational purposes, and would the definition include the workers only or their families as well? Where do U.S. citizens who do this type of work fit in, since they are subject to most of the same health problems as those who cross international boundaries to work in American fields
(2) What would be the funding source(s) for this national survey and what would be their research agenda for the study?
(3) What would be the funding level for the survey?
(4) Would the study be conducted by an established survey research organization or be a coordinated effort of a large number of more local groups (e.g., migrant health clinics, state health departments, county health departments)?

<u>First Stage</u>

Sampling unit = Small group (i.e., 3-6) of neighboring but not necessarily contiguous counties

Stratification = 3-5 migratory streams; concentration of MFWs according to the Department of Labor; other geographic predictors of health status

Selection method = PPS with estimated number of MFWs as the measure of size; estimate from the most recent census or other reliable source acceptable as long measures available at the county level

Issues: (1) How many counties and how large an area

should be covered by the counties grouped together as PSUs; larger groups tend to reduce the chances of double-counting especially mobile MFWs, while smaller groups would reduce the cost of field operations for sampling and interviewing

(2) Considerable cost savings could be realized by excluding counties with low concentrations of MFWs; the drawback to this would be a failure to achieve complete coverage since some MFWs would exist in the excluded areas

Second Stage

Sampling Unit = Enumeration districts (rural areas) or block groups (urban areas)

Stratification = Form two substrata: (1) with ED/BGs where the number of enumerated MFW households as of the last census exceeded some level (e.g., five) and (2) with all remaining ED/BGs

Selection = Select some constant number (b) of ED/BGs from substratum (1) with PPS using the number of enumerated MFWs in the last census as the size measure, and one ED/BG with equal probability from substratum (2)

Issues:

(1) b would be determined later on the basis of what is thought to be an optimum allocation of the sample within PSUs

(2) This stage relies heavily on the existence and availability of Bureau of Census ED/BG data on the number of MFWs and their households; although information on MFWs is available from the 1980 census and is planned for the 1990 census, these aggregated figures are not routinely produced and would require special arrangements with the Bureau; because of the Bureau's other reporting priorities, there might be some difficulty in getting the figures from the 1990 census very soon after the census is conducted; thus, anyone following this design for a MFW survey planned for the early 1990's could be faced by the following choices: (a) use 1980 figures on the number of MFWs for sampling purposes, (b) wait for the 1990

figures and face the possibility of delays in the survey, and (c) use some other size measure for PPS selection if the Bureau cannot easily produce the MFW figures

(3) This stage might conceivably be dropped if the MFW size measures for ED/BGs are not readily available.

Third Stage

Sampling unit = Migrant housing unit (i.e., a "housing unit" as defined by the Bureau of Census or an individual within a "group quarters" as defined by the Bureau)

Stratification = Type of migrant housing unit; form the following two substrata: (1) housing units and (2) individuals in group quarters; substratum (1) would tend to contain MFWs and their families, while substratum (2) would have a high percentage of single men living in houses and dormitories

Selection = Systematic sampling from a concatenated list of the migrant housing units in the two substrata

Issues:

(1) Frame construction will be a challenging and very labor-intensive task in this stage, relying heavily on the field worker's ability to find migrant camps and other habitations of MFWs; this will require help from local public and private groups providing service or assistance to MFWs (e.g., migrant health centers funded by the USPHS Office of Migrant Health, legal aid societies, church-affiliated relief organizations, migrant advocacy groups, local health departments, agricultural extension offices); information from several of these sources would have to be merged to produce the final frame

(2) Because of the amount of effort likely to be needed for frame construction, would it more cost=effective to have the field work for frame construction and the interviewing be done by the same person; traditionally this combination of activities has been avoided because of the potential for invalidated sample selection caused by field workers who

abandon the random methods in favor of sampling for the sake of convenience rather than science; of course frame could be sent to a supervisor or a central location for selection, but that may contribute to delays

Fourth Stage

Sampling unit = Individual migrant farm worker

Stratification = None currently planned, although the ordering by age for systematic sampling constitutes an implicit form of stratification;

Selection = Systematic sampling from an age-ordered roster of persons listed in the migrant housing unit

Issues:

(1) Explicit stratification will be needed if the decision is made to oversample certain segments of the MFW population (e.g., Haitians, females, legal MFWs, etc.); the calllenge then would be to find a suitable strategy for having the interviewer validly apply a randomized method with higher rates for the oversampled group(s)

(2) This stage of sampling will only be needed if, because of the type of data to be collected, it is necessary to select one person at random in each migrant housing unit (e.g., opinions are often contaminated if multiple interviews are conducted in a household, thus making a single interview per household desirable); in the event that one interview per household is not necessary, the logical choice for most types of dat is to interview all eligible members in each sample household

APPENDIX C

Paper on the Use of Time and Space Sampling for Surveys of Migrant Farm Workers

Design Strategies for Nonsedentary Populations

William D. Kalsbeek
Department of Biostatistics
University of North Carolina
Chapel Hill, NC 27514

ABSTRACT

One of the continuing frontiers of survey research is the design of samples for studies aimed at the highly mobile but less visible segments of society (e.g., the migrant farm worker, the teenage runaway, and the nomadic tribesman). Although sampling the places they visit has shown some promise for producing valid samples of these populations, this approach forces one to consider the complicating added dimension of time in choosing a suitable sampling design. Four designs for sampling jointly in space and time are described for the problem of estimating the population size (N), and the relative cost-efficiencies of an estimator of N are examined under these same designs. Empirical findings applicable to the problem of estimating the number of migrant farm workers in the United States reveal that stratified subsampling in time within spatial sampling units is generally preferred to unstratified subsampling within the same two-stage design framework, to a two-way design strategy in which the same sample in time is the same for each selected spatial unit, and to a design in which a simple with-replacement random sample of time periods in space is selected.

KEY WORDS: Sample design; Time sampling; Mobile populations; Population size estimation

1. INTRODUCTION

Migratory segments of the population in many countries are of considerable interest to the survey researcher. These segments (or "nonsedentary populations" as we shall call them in the sequel) occur in several different forms throughout the world. There are, for example, the migrant farm workers, the transients, the chronic homeless, and the teenage runaways in the United States; the European gypsies; the street people of Indonesia; and the nomadic tribesmen of East Africa.

This paper examines the problem of sampling nonsedentary populations. First I suggest that to characterize dynamic, mobile populations like these may require a design in which both space and time are sampled. Statistical and cost implications of four space-and-time design strategies are then formulated for the problem of estimating the size of the nonsedentary population. This is followed by an application of these results to the specific problem of sampling migrant farm workers in the United States. We conclude from this illustration that independently picking a stratified sample of days at each of a sample of migrant camps will be the most cost-efficient design strategy among those considered.

2. PRIOR SAMPLING STRATEGIES

Most prior surveys of nonsedentary populations have enumerated them as they appear at a sample of places where they reside during the study period. For example, Fernandez and Folkman (1975) and

Chi (1985) used labor camps as primary sampling units (PSUs) in multi-stage designs to sample migrant farm workers. In similarly structured designs, the Ministry of National Planning for Somalia (1981) designated watering points as PSUs to sample nomads, and Frankel (1986) used shelters, parks, streets and the like to sample the homeless in Chicago.

The final sample in each of these designs was chosen by identifying population members linked to the PSU at operationally convenient (but not randomly chosen) times during the data collection period. Random selection in these sampling designs is clearly limited to the spatial dimension, which implies that multiple frame linkages (i.e., multiplicity) exist during the study period and therefore must be accounted for in the estimation process. Moreover, the number and complexity of these linkages increase directly with the length of data gathering, thus making valid and efficient samples more difficult to obtain. Kalsbeek and Cross (1982) identified the sources and implications of multiplicity in sampling East African nomads, and Kalsbeek (1986) examined the properties of two alternative design strategies in this context. An extensive literature on multiplicity estimators, beginning with the work of Birnbaum and Sirken (1965) and Sirken (1970), addresses the matter of dealing with the multiple linkage issue.

In addition to the statistical implications mentioned above, extended periods of data collection may cause one difficulty in

trying to capture changes over time in the lives of nonsedentary individuals. One good example is the population of migrant farm workers whose size, composition and geographic distribution is known to change dramatically over a 12 month period (Johnston 1985). Workers move within well-known migratory streams to where the varying seasons among states provide a crop to be picked, occasionally returning to their homeland to visit family and friends. This mobility is important when the measures of interest in this population are tied to an individual's surroundings (e.g., health status, health care availability). A study conducted during the spring may paint a quite different picture than one done in the fall.

The designer of a one-time survey of a nonsedentary population is therefore faced with a fundamental dilemma. Should the reference period of the study be shortened to limit or avoid multiplicity, or should the study period be expanded to encompass all of the seasonal variations in behavior? It is the premise of this paper that one potentially useful resolution to this dilemma is to sample by means of a design in which both space and a fully expansive study period are sampled together so that variation along both dimensions can be examined through the sampling process while avoiding the problems of multiple linkage.

3. CONCEPTUAL FRAMEWORK AND PARAMETER

To characterize a nonsedentary population over time let us define a three-dimensional matrix, X_0, defined by the

cross-classification of the following:

1) An array made up of L sampling units constituting a frame for sampling spatially,

$$\mathbf{S} = (S_1, S_2, \ldots, S_i, \ldots, S_L)$$

plus the state, S_0, used to denote detachment from this frame (e.g., a migrant worker in transit between jobs or returning home to Mexico for a few months during the off-season);

2) An array of M time units (e.g., days) constituting the time period for the study and from which the sample in time is drawn,

$$\mathbf{T} = (T_1, T_2, \ldots, T_j, \ldots, T_M);$$

3) An array of N eligible members of some population being studied,

$$\mathbf{P} = (P_1, P_2, \ldots, P_k, \ldots, P_N).$$

Entries in $\mathbf{X_0}$ are denoted by X_{ijk}. It is assumed that during any time unit (T_j), every population member (P_k) will be linked to exactly one member of the frame (**S**) or to S_0. In addition, **P** is assumed to be an all-inclusive set of eligible population members during the study period defined by **T**.

Although totals, means, proportions and other more complex time-based parameters can be defined within $\mathbf{X_0}$, we will limit our attention here to the problem of estimating the population's size, N. To formulate this parameter, the entries of $\mathbf{X_0}$ are defined as count variables,

$$X_{ijk} = 1 \quad \text{if } P_k \text{ is linked to } S_i \text{ during } T_j, \text{ and}$$
$$= 0 \quad \text{if otherwise.}$$

The object as stated is to estimate

$$N = \sum_{i=0}^{L} \sum_{j}^{M} \sum_{k}^{N} X_{ijk}/M = L\,\bar{X}_0 ,$$

where $\sum_{j} \equiv \sum_{j=1}^{M}$, $\sum_{k} \equiv \sum_{k=1}^{N}$, and

$$\bar{X}_0 = \sum_{i=0}^{L} \sum_{j}^{M} \sum_{k}^{N} X_{ijk}/LM.$$

We note that designs for sampling in multiple dimensions have been considered in other areas of research. For example, Vos (1964) used two-dimensional sampling to estimate road transport usage over time. Although the precision of estimates was compared among several sampling designs, the cost implications of these designs were not considered in this comparison study. Sampling jointly in **S** and **T** also bears some resemblance to lattice or plane sampling as used in agricultural research (e.g., Bellhouse 1977; Iachan 1985). There, however, both dimensions are spatial and autocorrelation is expressible within the two-dimensional plane from which the sample is drawn. In the present setting the temporal is the only dimension within which any autocorrelation is likely to exist.

4. SAMPLING DESIGNS

The four sampling designs described below presume that population members cannot be sampled while in the detached state (S_0) and that the aggregate count of the number of population members linked to the i-th spatial sampling unit as of the j-th

temporal sampling unit,

$$X_{ij} = \sum_{k}^{N} X_{ijk} ,$$

can be determined for any combination of spatial and temporal sampling units. Moreover, to simplify formulations for components of the mean square error, with-replacement simple random sampling is assumed at each of the various selection steps of each design.

Unrestricted Random Sampling (U):

Each member of the unrestricted random sample (URS) of size 1*m (i.e., reads "1* times m," where "1*" denotes the lowercase Latin letter "el") is chosen from the LM members of **S** by **T** by picking one spatial sampling unit at random and then choosing a temporal sampling unit at random to go with it. Subsequent selections are made without regard to prior selections (i.e, with replacement). Let "URS(1*m of LM)" be a shorthand way to describe this design, which has the advantage of avoiding the negative statistical implications of sampling time clusters whose intracluster homogeneity may be high.

Unstratified 2-Stage (2S):

A URS(1* of L) spatial sampling units is chosen as the primary sample; a URS(m of M) temporal sampling units is independently selected within each sample PSU. This design limits to 1* the number of spatial sampling units that must be visited during data collection.

Substratified 2-Stage (2SS):

A URS(1* of L) spatial sampling units is chosen in the first stage; a proportionate stratified URS of size m is chosen from H_i strata formed in the i-th PSU. This design compensates for the losses due to cluster sampling by gains due to stratification in the second stage of sampling.

Unstratified 2-Way (2W):

The same URS(m of M) sample of temporal sampling units is used for each member of a URS(1* of L) sample of spatial sampling units. This design has the potentially useful feature of having data gathering at the same time points in all selected spatial units. It represents an effort to coordinate the timing of selected time units among spatial units, which in some instances might be useful (e.g., when special preparations are needed for collecting data at each time point).

5. ESTIMATION

Since each of the four designs yields a sample size of 1*m as well as equal selection probabilities for each of LM cells in S, a common (though biased) estimator of N would be,

$$\hat{N} = L \left[\sum_{i}^{1*} \sum_{j}^{m} X_{ij} / 1*m \right] = L \bar{x} ,$$

where $\sum_{i}^{1*} \equiv \sum_{i=1}^{1*}$ and $\sum_{j}^{m} \equiv \sum_{j=1}^{m}$.

6. RELATIVE BIAS OF $\hat{N}$

The bias of $\hat{N}$ arising from coverage error in the frame S is the

same for each design and can be obtained by noting that (over all possible samples in space and time)

$$E(\bar{x}) = \sum_{i=1}^{L} \sum_{j}^{M} X_{ij}/LM ,$$

so that the relative bias of $\hat{N}$ is,

$$\text{Rel-Bias}(\hat{N}) \equiv \text{Bias}(\hat{N})/N = - \sum_{j}^{M} X_{0j}/NM . \qquad (6.1)$$

In Eq.(6.1) we see our first evidence of how the nature of mobility in a nonsedentary population has an impact on the properties of estimators used in conjunction with these four designs. All else constant, we note that a population which tends to be detached from the spatial sampling frame frequently, for longer intervals, and in greater numbers will lead to greater underestimation than one where periods of detachment are less frequent and lengthy. Thus $\hat{N}$ will be a relatively larger underestimate when used to estimate the number of nomads in East Africa, with their collectively much lower dependency on watering points during significant and unpredictable periods of the year, than when used to determine the number of migrant farm workers whose periods of travel between jobs are relatively short and returns to their home countries infrequent.

7. VARIANCE OF $\hat{N}$ AND DESIGN EFFECTS

To formulate the variances of $\hat{N}$ under the four designs, we must first define the following measures of variance:

$$\sigma^2_{BL} = \sum_{i=1}^{L} (\bar{X}_i - \bar{X})^2/L \ , \qquad \text{(Between-Space)}$$

$$\sigma^2_{WL} = \sum_{i=1}^{L} \sum_{j}^{M} (X_{ij} - \bar{X}_i)^2/LM \qquad \text{(Within-Space)}$$

$$\sigma^2_{BM} = \sum_{j}^{M} (X_j - \bar{X})^2/M \qquad \text{(Between-Time)}$$

$$\sigma^2_{WM} = \sum_{i=1}^{L} \sum_{j}^{M} (X_{ij} - \bar{X}_j)^2/LM \qquad \text{(Within-Time)}$$

$$\sigma^2 = \sum_{i=1}^{L} \sum_{j}^{M} (X_{ij} - \bar{X})^2/LM \ , \qquad \text{(Overall)}$$

$$= \sigma^2_{BL} + \sigma^2_{WL} = \sigma^2_{BM} + \sigma^2_{WM} \ ,$$

$$\delta_L = \sigma^2_{BL}/\sigma^2 \ , \qquad \text{(Relative Homogenity within Spatial Clusters)}$$

$$\delta_M = \sigma^2_{BM}/\sigma^2 \ , \qquad \text{(Relative Homogenity within Temporal Clusters)}$$

where $\bar{X}_i = \sum_{j}^{M} X_{ij}/M$ and $\bar{X}_j = \sum_{i=1}^{L} X_{ij}/L$. Note that the relative homogeneity measures are comparable but not equivalent to the usual measures of intracluster correlation in that $0 \le \delta_L \le 1$ and $0 \le \delta_M \le 1$.

URS(1*m of LM)

The variance of $\hat{N}$, from the well-known expression for $Var(\bar{x})$, is

$$Var_U(\hat{N}) = L^2\sigma^2/1*m \ . \qquad (7.1)$$

Unstratified 2-Stage

Once again the variance of $\bar{x}$ is known from the standard 2-stage framework, here with spatial sampling units as PSUs and temporal sampling units as secondary sampling units independently selected in each sample PSU.

The design effect for $\hat{N}$ given this 2-stage design, determined as its variance, relative to the variance of the estimator of N for a URS(1*m of LM), will be

$$DEFF_{2S}(\hat{N}) = 1 + \delta_L(m-1) \ . \qquad (7.2)$$

Substratified 2-Stage

When a proportionate stratified URS of size m replaces the URS(m of M) in the second stage of the 2-stage design above, the variance of $\bar{x}$ can be expressed as,

$$Var(\bar{x}) = \sigma^2_{BL}/1^* + \{\sum_{i=1}^{L} (1-\delta_{iH})\sigma^2_i/L\}/1^*m \ ,$$

where $\delta_{iH} = \sigma^2_{iBH}/\sigma^2_i$ measures the effectiveness of sub-stratification with H_i substrata in the i-th spatial sampling unit,

$$\sigma^2_{iBH} = \sum_{h=1}^{H_i} W_{ih}(\bar{X}_{ih}-\bar{X}_i)^2 \ ,$$

is the between-substratum variance in the i-th PSU,

$$\sigma^2_i = \sum_{h=1}^{H_i} \sum_{j}^{M_{ih}} (X_{ihj}-\bar{X}_i)^2/M \ ,$$

is the total within-cluster variance for the i-th PSU, $W_{ih}=M_{ih}/M$ is the proportion of time units in the i-th PSU that fall in its

h-th substratum, $\bar{X}_{ih} = \sum_{j}^{M_{ih}} X_{ihj}/M_{ih}$, and

$$\bar{X}_i = \sum_{h=1}^{H_i} \sum_{j}^{M_{ih}} X_{ihj}/M = \sum_{h=1}^{H_i} W_{ih}\bar{X}_{ih} .$$

When the efficiency of substratification and σ_i^2 are uncorrelated among PSUs, then one can express the design effect under the substratified 2-stage design from the $\mathrm{Var}(\bar{x})$ given above as

$$\mathrm{DEFF}_{2SS}(\hat{N}) = 1 + \delta_L(m-1) - \bar{\delta}_H(1-\delta_L) , \qquad (7.3)$$

where $\bar{\delta}_H = \sum_{i=1}^{L} \delta_{iH}/L$.

Unstratified 2-Way

The variance here is found by reformulating the overall sample mean of the X_{ij}'s as

$$\bar{x} = \sum_{i=1}^{L} \sum_{j}^{M} \theta_i\theta_j X_{ij}/l^*m ,$$

where θ_i and θ_j are, respectively, the number of times that the i-th spatial sampling unit and the j-th temporal sampling unit are chosen. From this we obtain

$$\mathrm{Var}(\bar{x}) = \sigma_{WL}^2/m + \sigma_{WM}^2/l^* - \{l^*+m-1\}\sigma_{LM}/l^*m ,$$

where

$$\sigma_{LM} = \sum_{i=1}^{L} \sum_{j}^{M} (X_{ij}-\bar{X}_i)(X_{ij}-\bar{X}_j)$$

is the space-time interaction for the X_{ij}'s. Finally, the design

effect for the 2-way design as used to estimate N can be shown as

$$\hat{DEFF}_{2W}(N) = 1 + \delta_L(m-1) + \delta_M(1^*-1). \qquad (7.4)$$

We note from Eqs.(7.2)-(7.4) that, when considering variances alone, the substratified 2-stage design is always preferable to an unstratified 2-stage design which, in turn, is always preferable to the unstratified 2-way design. The ranking of URS(1*m of LM), relative to the substratified 2-stage design, depends on values of δ_L, $\bar{\delta}_H$, and m. More specifically, $DEFF_{2SS} < DEFF_U$ when

$$\bar{\delta}_H > \delta_L(m-1)/(1-\delta_L) \; . \qquad (7.5)$$

8. SURVEY COSTS

Having assessed the primary statistical implications of the four designs, let us now turn our attention to how each design would affect the cost of the survey operation. To do so, we need a model to express these costs. One simple formulation of costs incurred under these designs is the following:

$$C = C_o + \sum_{i=1}^{L} \varphi_i C_{Li} + \sum_{i=1}^{L} \sum_{j}^{M} \varphi_{ij} C_{Mij} \; , \qquad (8.1)$$

where C_o denotes fixed costs (e.g., instrument development, administration, reporting), C_{Li} represents those costs that are particular to the i-th spatial sampling unit (e.g., solicitation, set-up, certain sampling activities), C_{Mij} denotes the cost of survey activity (e.g., data collection and processing) tied to the

j-th time unit in the i-th spatial unit,

$\varphi_i = 1$ if ≥1 time units are chosen in the i-th spatial unit;
$= 0$ if otherwise, and

$\varphi_{ij} = 1$ if the j-th time unit in the i-th spatial unit is chosen ≥1 times;
$= 0$ if otherwise.

Allowing $1-\{1-1/L\}^{1*} \approx 1*/L$ and $1-\{1-1/M\}^{m} \approx m/M$, the expected cost of the survey under the URS(1*m of LM) design will be

$$E_U(C) \approx C_o + 1*m(\bar{C}_L + \bar{C}_M) , \tag{8.2}$$

where $\bar{C}_L = \sum_{i=1}^{L} C_{Li}/L$ and $\bar{C}_M = \sum_{i=1}^{L}\sum_{j}^{M} C_{Li}/LM$.

Allowing the same approximations for 1*/L and m/M, once again, the expected costs for the other three designs will be

$$E_{2S}(C) \approx E_{2SS}(C) \approx E_{2W}(C) \approx C_o + 1*\bar{C}_L + 1*m\bar{C}_M . \tag{8.3}$$

We note from Eqs. (8.2) and (8.3) that the URS(1*m of LM) design will have non-fixed costs (i.e., excluding C_o) that exceed comparable costs under the other designs by a factor of

$$\text{REL-COST} = 1 + R_{LM}(m-1)/(R_{LM}+m) . \tag{8.4}$$

where $R_{LM}=\bar{C}_L/\bar{C}_M$ is the ratio of average spatial to temporal unit costs.

9. COST-EFFICIENCY

The overall measure of effectiveness adopted for use in assessing any design (*) in terms of its joint statistical and

fiscal impact is

$$CEFF_{\star}(\hat{N}) = \frac{\{Var_{\star}(\hat{N})\}^{-1}}{E_{\star}(C)-C_o},$$

whose numerator reflects the statistical precision obtainable from the design and denominator accounts for the non-fixed, or variable, component of survey costs over which the designer has some control.

Because our real interest is in comparing the cost-effectiveness of the four designs when l^* and m are the same, we choose to examine CEFF for any given design relative to the measure of CEFF for the URS(l^*m of LM) design; i.e., we use

$$\begin{aligned} RCEFF_{\star}(\hat{N}) &= CEFF_{\star}(\hat{N})/CEFF_{U}(\hat{N}) \\ &= REL\text{-}COST/DEFF_{\star}(\hat{N}) \end{aligned} \qquad (9.1)$$

as the final basis for comparison among the four designs.

10. ILLUSTRATION: MIGRANT FARM WORKERS

We illustrate our findings by considering the feasibility of the four designs for estimating the number of migrant farm workers in the United States during a one year period. In this setting the penultimate spatial sampling units are presumed to be migrant camps, although technically other residential areas inhabited by high concentrations of migrant workers would be included to improve sample coverage. The temporal sampling units are individual days, and each value of X_{ij}, the headcount of migrant workers on a specific day at a specific camp, is obtained by a

visit to the camp on that day. Days must be selected from the both the growing season and from the off-season, when enrollment in the camps is much lower and limited to those migrants with more permanent work in the area.

Because there exists little direct information on the size of other key design parameters, we must rely on quasi-empirical evidence to establish values for δ_L, δ_M, R_{LM}, $\bar{\delta}_H$, m, and l^*. One key piece of evidence has to do with the pattern (not distribution) of the daily census of each camp (i.e., X_{ij}) from the first to last day of the year, since from this pattern one can obtain σ_i^2, σ_{ih}^2, and ultimately values for $\bar{\delta}_H$ and δ_L. Standardized means and variances for two commonly observed patterns are presented in Figure 1.

The "partial square pattern" presumes that the camp is occupied at full capacity for 100c percent of the growing season, which occurs for 100α percent of the year. This pattern is seen in locations where the work is done by crews, which arrive together early in the season and then move on together to the next job near season's end. Substratification here assumes that only the starting and ending dates of the growing season are known.

The "trapezoid pattern" presumes a peak season occurring for 100β percent of the growing season and a head count that gradually increases to full capacity at the start of the peak growing season and then diminishes in like manner toward the end. This pattern is common in "home-base" states like California, Texas, and

Florida where migrants may return to semi-permanent residences around the peak of the growing season in those states. These residences serve as bases of movement to other states for their growing seasons. Starting and ending dates for the growing season, as well as its peak, are once again assumed known in defining substrata.

Values for δ_L:

The following two key assumptions are made in arriving at the values of δ_L that are used in our illustration:

(1) All camps follow a partial square pattern with ε=1 and with a peak enrollment per camp (at its average) of 75 persons and a growing season (at its average) of 9 months (Johnston, 1985); and

(2) The distribution of the peak enrollment among all camps is asymmetrically triangular with a range of from 0 to 200.

These assumptions imply δ_L = 0.51, which led to the use of δ_L = 0.4, 0.5 and 0.6 in this illustration.

Values for δ_M:

Here it is thought that δ_M must be quite small, since variation in the average aggregate enrollment in camps ($\bar{X}_j$) over time is likely to be small. This reasoning implicitly assumes that the total number of migrants in the detached state (S_0) will not vary much from one day to the next in a year. The values δ_M = 0.01, 0.03 and 0.05 are therefore used.

Values of $\bar{\delta}_H$:

The effectiveness of substratification in an individual camp (δ_{iH}) will depend on the camp's enrollment pattern during the year and on the number and definition of substrata. For the partial square pattern where there are two substrata as indicated in Figure 1,

$$\delta_{iH} = 1 - \{(1-\varepsilon)/(1-\alpha\varepsilon)\} . \qquad (10.1)$$

In camps where enrollment follows a trapezoidal pattern and three substrata are formed, one covering the peak period, a second the off-peak portion of the growing season, and the third the rest of the year,

$$\delta_{iH} = \frac{3\{(3\beta+1)-\alpha(\beta+1)^2\}}{4(2\beta+1)-3\alpha(\beta+1)^2} . \qquad (10.2)$$

Table 1 presents δ_{iH} for various values of α, β and ε. Since the majority of camps following the partial square pattern will be in non-"home-base" states where most growing seasons are 6-9 months long, we see that δ_{iH} is likely to exceed 0.5 there. Camps in home-base states have seasons nearly year-round which would imply δ_{iH} between 0.3 and 0.5. Finally, assuming that 60-80 percent of camps follow the partial square pattern, it seems plausible that the overall measure of the substratification efficiency ($\bar{\delta}_H$) might comfortably be encompassed by the values 0.5, 0.7 and 0.9.

Values of R_{LM}:

Thinking of survey cost in person-days of effort, and assuming that it takes roughly one person-day of effort to visit a camp on a selected day and to process the measure of $\bar{X}_{ij}$ through analysis, then $R_{LM}=\bar{C}_L$, the average number of person-days needed to add a camp to the sample.

The size of $\bar{C}_L$ will depend on several things. First, since the sample of camps is likely to be chosen through some multi-stage process, part of the cost per camp will depend on the amount of effort expended in developing lists of existing migrant camps to be used as sampling frames. Prior experience has shown that frame construction can be very costly if the object is to achieve high coverage rates for these frames (C. Littlefield, personal communication, September 2, 1987). A second determinant of $\bar{C}_L$, related to the first, is the number of such frames to be constructed, which would depend on the allocation of the spatial sample among the stages identified for selection. Clearly, $\bar{C}_L$ would vary directly as the number of such frames to construct. Finally, $\bar{C}_L$ would be directly affected by the effort expended in training, supervision, quality control and the like.

Given that R_{LM} could be high or low, depending on the priorities of the study, the measures used in our illustration reflect this uncertainty and thus values of R_{LM} ranging from five to 50 are used.

Values of m:

Values for m in the illustration were based on the cost-times-variance optimum values of m that would arise from the variances of the three non-URS designs, using the model for the variable components of their corresponding expected costs as presented in Eq. (8.3). Optimum values for these designs can be determined for the unstratified 2-stage design (widely used) as

$$m_{2S}^{(opt)} \doteq [(1-\delta_L)R_{LM}/\delta_L]^{1/2} , \tag{10.3}$$

for the substratified 2-stage design as

$$m_{2SS}^{(opt)} \doteq [(1-\bar{\delta}_H)(1-\delta_L)R_{LM}/\delta_L]^{1/2} , \tag{10.4}$$

and for the unstratified 2-way design as

$$m_{2W}^{(opt)} \doteq [\{(1-\delta_L)+ \delta_M(1^*-1)\}R_{LM}/\delta_L]^{1/2} . \tag{10.5}$$

Optimum values of m for both versions of the 2-stage design are presented in Table 2, where we note that $4 \leq m \leq 8$ is generally optimum for the unstratified 2-stage design and that $2 \leq m \leq 4$ is often preferable for the substratified 2-stage design. Optimum values of m, subject to the availability of 5,000 person-days of variable costs and C_M = one person-day, are presented for the unstratified 2-way design in Table 3. There one notes that $10 \leq m \leq 18$ generally covers the range of optimum values.

The values, m = 2, 6 and 10, used in later computations are intended to represent the values most likely to be effective under each of the designs other than URS(1*m of LM).

Findings on Relative Cost-Efficiency:

Table 4 presents values of the cost-efficiency of each design relative to URS (i.e., RCEFF). Values less than 1 indicate that the URS(1*m of LM) design is more cost-efficient by our criteria, while those greater than 1 point to the referent design being preferable to URS. Relative superiority and inferiority among the non-URS designs can also be gauged using these entries.

Several potentially useful findings can be inferred from Table 4. First, the substratified 2-stage design is generally the most cost-efficient among the four designs considered. Its preference is due mainly to its lower variance than the other non-URS designs, with which it shares notably lower non-fixed costs than the URS design. As expected, its strongest showing overall occurs when m is relatively small. The unstratified 2-stage design is preferable to URS in the majority of instances, implying that substratification is not neccessarily needed to counteract the substantial variance increase due to cluster sampling with large δ_L. Second, the two 2-stage designs are most similar in preference and substantially superior to the unstratified 2-way design when R_{LM} is low. The overall last-place showing of the 2-way design is largely due to the size of 1* which amplifies its design effect, even with relatively small values of δ_M. The 2-way design is most competitive with the URS design when both R_{LM} and m are relatively large. The superiority of the 2-way design in this case is atttributable to its relatively moderate design effect

combined with its substantial cost savings over the URS design. Finally, the only notable instance where the URS design does well is when R_{LM} is lowest and m is highest among observed values.

Discussion:

Findings in Table 4 generally portray the substratified 2-stage design as the one of choice among the four considered when estimating population size, given cost models where equal variable cost components for the non-URS designs are much lower than comparable costs for the URS design. One must then wonder if and how these findings might be altered for other related design settings. For example, how might the comparison of the two 2-stage designs been altered if the cost of stratification had been allowed to increase the variable costs of the substratified 2-stage design? Findings not presented revealed that the stratified design is still generally preferred over the unstratified design. Another facet of the assumptions of this study that must be examined is the effect of added complexity in the design (e.g., multi-stage cluster sampling) used to choose the sample of 1* spatial sampling units. Here the implications are less clear-cut, although we suspect that changes in the absolute sizes of RCEFF are not likely to be great since each design would experience similar increases in both variance and cost. These claims are of course conjectural and must be substantiated by empirical data to allow one to better choose among the options given to sample migrant seasonal farm workers and other nonsedentary populations.

REFERENCES

Bellhouse, D.R. (1977), "Optimal Designs for Sampling in Two Dimensions," Biometrika, 64, 606-611.

Birnbaum, Z.W. and Sirken, M.G. (1965), "Design of Sample Surveys to Estimate the Prevalence of Rare Diseases: Three Unbiased Estimates," Vital and Health Statistics, Series 2 - Number 11, National Center for Health Statistics, U.S. Government Printing Office, Washington, DC.

Chi, P.S.K. (1985), "A Note on Sampling Migrant Farm Workers," Rural Sociologist, 5, 3, 158-162.

Fernandez, R.R. and Folkman, D.V. (1975), "Survey of Vocational Educational Needs of Migrants in Wisconsin," University of Wisconsin, Extension Center for Community Leadership Development, Madison, WI.

Frankel, M. R. (1986), "A Probability Sample of the Homeless Population of Chicago," Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 176-177.

Iachan, R. (1985), "Plane Sampling," Statistics and Probability Letters, 3, 151-159.

Johnston, H.L. (1985), Health for the Nation's Harvesters, National Migrant Workers Council, Inc., Farmington Hills, MI.

Kalsbeek, W.D. and Cross, A.R. (1982), "Problems in Sampling Nomadic Populations," Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 398-402.

Kalsbeek, W.D. (1986), "Nomad Sampling: An Analytic Study of Alternative Design Strategies," Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 164-169.

Kish, L. (1965), Survey Sampling, John Wiley and Sons, New York.

Ministry of National Planning (1981), Report on the Demograghic Survey of Banadir, Bay, and Lower Shebelle Regions of Somalia. 1980-81, Mogadishu. Also found in Report Summary Series No. 4, International Population Laboratories, Unversity of North Carolina, Chapel Hill, NC.

Sirken, M.G. (1970), "Household Surveys With Multiplicity," Journal of the American Statistical Association, 65, 257-266.

Vos, J.W.E. (1964), "Sampling in Space and Time," Review of the International Statistical Institute, 32, 226-241.

TABLE 1: VALUES OF δ_{1H} FOR THE PARTIAL SQUARE AND TRAPEZOID PATTERNS IN SAMPLING MIGRANT FARM WORKERS

PROPORTION OF SEASON IN RESIDENCE (ε)	PROPORTION OF THE YEAR SPANNED BY THE GROWING SEASON (α) =	NUMBER OF MONTHS IN THE GROWING SEASON									
		3	4	5	6	7	8	9	10	11	12
		0.25	0.33	0.42	0.50	0.58	0.67	0.75	0.83	0.92	1.00
PARTIAL SQUARE PATTERN/ 2-STRATUM SEASON CONFIGURATION:											
0.75		0.69	0.67	0.64	0.60	0.56	0.50	0.43	0.33	0.20	0.00
0.80		0.75	0.73	0.70	0.67	0.63	0.57	0.50	0.40	0.25	0.00
0.85		0.81	0.79	0.77	0.74	0.70	0.65	0.59	0.49	0.32	0.00
0.90		0.87	0.86	0.84	0.82	0.79	0.75	0.69	0.60	0.43	0.00
0.95		0.93	0.93	0.92	0.90	0.89	0.86	0.83	0.76	0.61	0.00
1.00		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	----
TRAPEZOID PATTERN/3-STRATUM CONFIGURATION ASSUMING A PEAK OF 2 MONTHS IN ALL STATES:											
----		0.95	0.91	0.87	0.83	0.79	0.74	0.68	0.60	0.49	0.33
	β =	0.67	0.50	0.40	0.33	0.29	0.25	0.22	0.20	0.18	0.17

NOTE: δ_{1H}, RANGING BETWEEN 0 AND 1, MEASURES THE EFFECTIVENESS OF SUBSTRATIFICATION in the i-TH CAMP.

TABLE 2: OPTIMUM NUMBER OF DAYS (m) TO SELECT PER CAMP IN SAMPLING MIGRANT FARM WORKERS BY A 2-STAGE DESIGN WITH AND WITHOUT SUBSTRATIFICATION

RATIO: CAMP UNIT COST DIVIDED BY TIME UNIT COST (R_{LH})		UNSTRATIFIED TIME SAMPLING			STRATIFIED TIME SAMPLING								
	$\bar{\delta}_H$ =	0.0			0.5			0.7			0.9		
	δ_L =	0.4	0.5	0.6	0.4	0.5	0.6	0.4	0.5	0.6	0.4	0.5	0.6
5		3	2	2	2	2	1	2	1	1	1	1	1
10		4	3	3	3	2	2	2	2	1	1	1	1
15		5	4	3	3	3	2	3	2	2	1	1	1
20		5	4	4	4	3	3	3	2	2	2	1	1
25		6	5	4	4	4	3	3	3	2	2	2	1
30		7	5	4	5	4	3	4	3	2	2	2	1
35		7	6	5	5	4	3	4	3	3	2	2	2
40		8	6	5	5	4	4	4	3	3	2	2	2
45		8	7	5	6	5	4	5	4	3	3	2	2
50		9	7	6	6	5	4	5	4	3	3	2	2

NOTE: $\bar{\delta}_H$ MEASURES THE EFFECTIVENESS OF SUBSTRATIFICATION IN TIME; δ_L MEASURES THE WITHIN-CAMP HOMOGENEITY AMONG DAYS.

TABLE 3: OPTIMUM NUMBER OF DAYS (m) TO SELECT PER CAMP IN SAMPLING MIGRANT FARM WORKERS BY AN UNSTRATIFIED 2-WAY DESIGN IN A SURVEY WITH 5,000 PERSON-DAYS AVAILABLE FOR NON-FIXED COST ACTIVITY

RATIO: CAMP UNIT COST DIVIDED BY TIME UNIT COST (R_{LM})	δ_M = 0.01			0.03			0.05		
δ_L =	0.4	0.5	0.6	0.4	0.5	0.6	0.4	0.5	0.6
5	8	7	6	11	10	9	13	12	11
10	9	8	7	13	12	11	16	15	14
15	10	9	8	15	13	12	18	16	15
20	11	9	8	15	14	13	19	17	16
25	11	10	9	16	15	13	20	18	16
30	12	10	9	17	15	14	20	18	17
35	12	11	9	17	16	14	21	19	17
40	12	11	10	18	16	14	21	19	18
45	13	11	10	18	16	15	22	20	18
50	13	11	10	19	17	15	22	20	18

NOTE: (1) δ_M MEASURES WITHIN-DAY HOMOGENEITY AMONG CAMPS; δ_L MEASURES WITHIN-CAMP HOMOGENEITY AMONG DAYS;

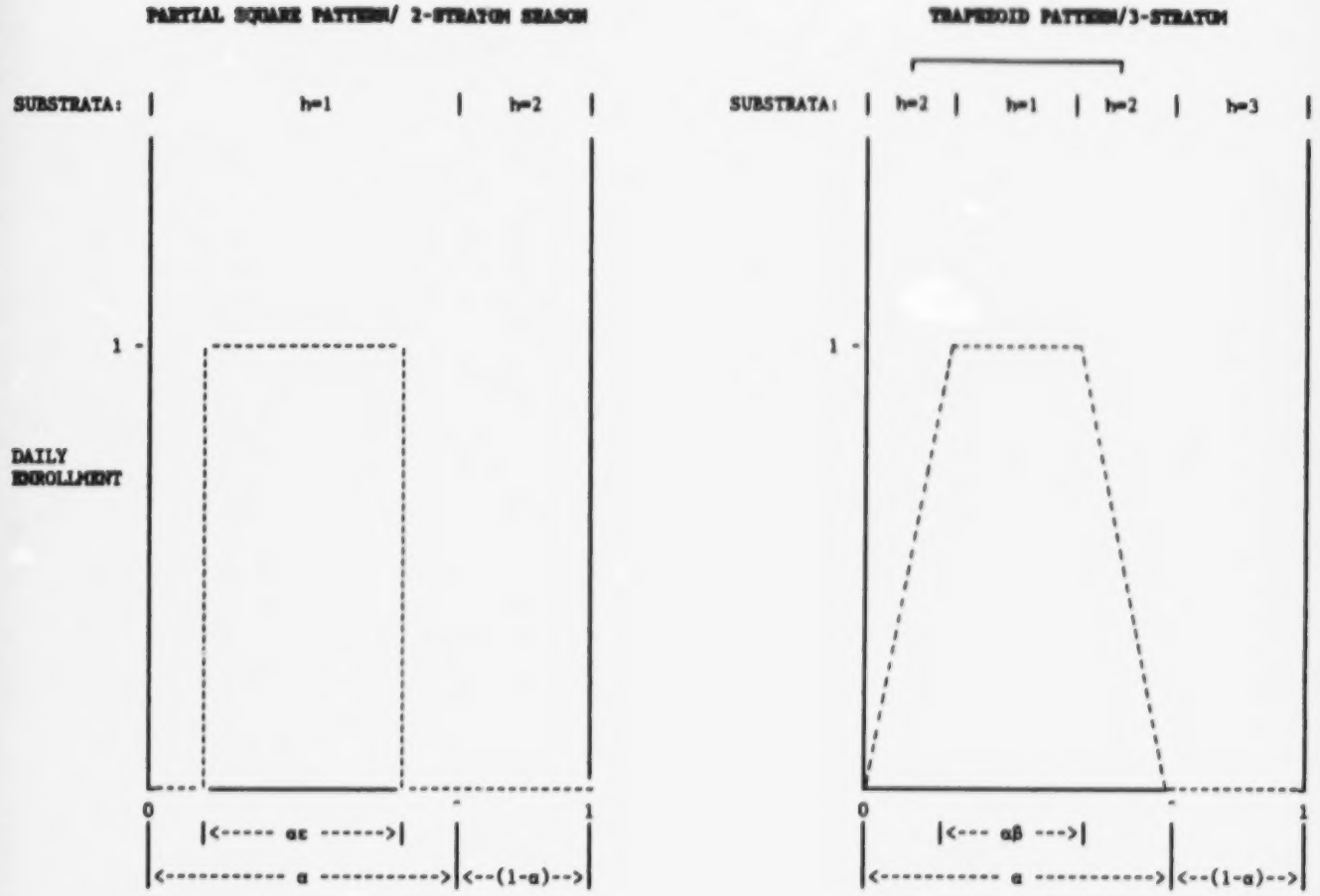
(2) THE AVERAGE UNIT COST AMONG ALL SELECTED DAYS IS ONE PERSON-DAY OF SALARY, I.E., $\bar{C}_M$=1.

TABLE 4: RELATIVE COST-EFFICIENCY (RCEFF) FOR 2-STAGE AND 2-WAY DESIGNS COMPARED TO A URS(1*m of 1M) DESIGN

SAMPLE SIZES		RATIO OF UNIT COST	WITHIN-CAMP HOMOGENEITY	SUBSTRATIFIED 2-STAGE $\bar{\delta}_H$			UNSTRATIFIED 2-STAGE	UNSTRATIFIED 2-WAY δ_H		
DAYS (m)	CAMPS (1*)	(R_{LM})	(δ_L)	0.5	0.7	0.9	---	0.01	0.03	0.05
2	714	5	0.4	1.56	1.75	1.99	1.22	0.20	0.08	0.05
			0.5	1.37	1.49	1.63	1.14	0.20	0.07	0.05
			0.6	1.22	1.30	1.38	1.07	0.20	0.07	0.05
2	185	25	0.4	1.75	1.97	2.24	1.38	0.59	0.28	0.18
			0.5	1.54	1.67	1.83	1.28	0.58	0.27	0.18
			0.6	1.38	1.46	1.55	1.20	0.56	0.27	0.18
2	96	50	0.4	1.78	2.00	2.28	1.40	0.83	0.46	0.32
			0.5	1.57	1.71	1.87	1.31	0.80	0.45	0.31
			0.6	1.40	1.49	1.58	1.23	0.77	0.44	0.31
6	455	5	0.4	1.21	1.27	1.33	1.09	0.43	0.20	0.13
			0.5	1.01	1.04	1.07	0.94	0.41	0.19	0.13
			0.6	0.86	0.88	0.90	0.82	0.38	0.19	0.12
6	161	25	0.4	1.86	1.95	2.05	1.68	1.09	0.64	0.46
			0.5	1.55	1.60	1.65	1.44	0.99	0.61	0.44
			0.6	1.32	1.35	1.38	1.26	0.90	0.57	0.42
6	89	50	0.4	2.02	2.12	2.22	1.82	1.41	0.97	0.74
			0.5	1.68	1.73	1.79	1.56	1.25	0.89	0.69
			0.6	1.44	1.47	1.50	1.37	1.12	0.82	0.65
10	333	5	0.4	0.93	0.96	0.99	0.87	0.50	0.27	0.19
			0.5	0.76	0.78	0.79	0.73	0.45	0.26	0.18
			0.6	0.65	0.65	0.66	0.63	0.41	0.24	0.17
10	143	25	0.4	1.73	1.78	1.83	1.61	1.23	0.84	0.64
			0.5	1.41	1.44	1.47	1.35	1.07	0.76	0.59
			0.6	1.20	1.21	1.23	1.16	0.95	0.70	0.55
10	83	50	0.4	1.98	2.03	2.09	1.85	1.57	1.20	0.98
			0.5	1.62	1.65	1.68	1.55	1.34	1.07	0.88
			0.6	1.37	1.39	1.41	1.33	1.18	0.96	0.81

NOTE: COMPUTATIONS FOR 1* ASSUME A TOTAL OF 5,000 PERSON-DAYS IN NON-FIXED COSTS FOR THE SURVEY.

FIGURE 1: STANDARIZED MEANS AND VARIANCES OF DAILY ENROLLMENT PATTERNS FOR MIGRANT CAMPS



MEAN = $\alpha\varepsilon$

VARIANCE = $\alpha\varepsilon(1-\alpha\varepsilon)$

MEAN = $\alpha(1+\beta)/2$

VARIANCE = $\alpha[4(2\beta+1)-3\alpha(1+\beta)^2]/12$

NOTES:

(1) STANDARDIZATION IS TO UNIT LENGTH AND UNIT MAXIMUM ENROLLMENT. MEANS FOR A CAMP WITH A MAXIMUM ENROLLMENT OF X* CAN BE OBTAINED BY MULTIPLYING THE STANDARDIZED MEAN BY X*. VARIANCES CAN BE OBTAINED BY MULTIPLYING BY X^{*2}.

(2) MEANS AND VARIANCES ARE UNAFFECTED BY DEPARTURES FROM SYMMETRY OF THE INTERVAL OF PEAK ENROLLMENT WITHIN THE GROWING SEASON.